

D	b	181	GGCCACCGATAGAGACTGTGCACCGAGGCGCAGAGCCAACTTCTTCCGCGAGACCTTGACC	240
Q	y	241	TTCCCCAGAGGAGCCGCGAGTTCCCGAGGACAGAACCGCGCCAAACAGCCCCACC	300
D	b	241	TTCCCCAGAGGAGCCGCGAGTTCCCGAGGACAGAACCGCGCCAAACAGCCCCACC	300
Q	y	301	AGCCGCGAGCTGCAGGTGTGGCGCGCCAAACCCCTCGACGCGAGGCGCGCTCGAGCGCCAG	360
D	b	301	AGCCGCGAGCTGCAGGTGTGGCGCGCCAAACCCCTCGACGCGAGGCGCGCTCGAGCGCCAG	360
Q	y	361	GGCACCTCTGAACTTCCCCAGATCAACCTGTGTGCAGCGCCCCCTGTGTAGCATCAAGTGT	420
D	b	361	GGCACCTCTGAACTTCCCCAGATCAACCTGTGTGCAGCGCCCCCTGTGTAGCATCAAGTGT	420
Q	y	421	GGCGGCGCAGATCAAGAGGCGCCCTGCTGACAACCGAGCGCGACACACCGTGTGAGAGAG	480
D	b	421	GGCGGCGCAGATCAAGAGGCGCCCTGCTGACAACCGAGCGCGACACACCGTGTGAGAGAG	480
Q	y	481	ATGAGACTGTGCCCCGCGCAATGTGAAGCCCAAGATGATGTGGCGCATGTGGCGCTTCAATCAAG	540
D	b	481	ATGAGACTGTGCCCCGCGCAATGTGAAGCCCAAGATGATGTGGCGCATGTGGCGCTTCAATCAAG	540
Q	y	541	GTGGCGCAGTACGACCGAGATCCCTGATCGAGATTTGGCGCAAGAGGCGCATGTGCACCGTGTG	600
D	b	541	GTGGCGCAGTACGACCGAGATCCCTGATCGAGATTTGGCGCAAGAGGCGCATGTGCACCGTGTG	600
Q	y	601	CTGATCGGCGCCACCCCGCTGAAACATCATCTGCGCGCAACATGTCTGACCCACGTGGGCTGC	660
D	b	601	CTGATCGGCGCCACCCCGCTGAAACATCATCTGCGCGCAACATGTCTGACCCACGTGGGCTGC	660
Q	y	661	ACCTTGAACTTCCCATCAAGCCCATCGAGACCTGTGCGCGGTGAAGCTGAAAGCCCGCGCATG	720
D	b	661	ACCTTGAACTTCCCATCAAGCCCATCGAGACCTGTGCGCGGTGAAGCTGAAAGCCCGCGCATG	720
Q	y	721	GACGCGCCCAAGGTGAAGAGATGTGCGCCCTGACCGAGAGAAAGATTCAAAGCGCTGACCGCC	780
D	b	721	GACGCGCCCAAGGTGAAGAGATGTGCGCCCTGACCGAGAGAAAGATTCAAAGCGCTGACCGCC	780
Q	y	781	ATCTGCGAGAGATGTGAGAGAGGCGCAAGATCAACAAAGATGTGCGCCCGAGAACCCCTTAC	840
D	b	781	ATCTGCGAGAGATGTGAGAGAGGCGCAAGATCAACAAAGATGTGCGCCCGAGAACCCCTTAC	840
Q	y	841	AAACAACCCCGCTGTTCGCGCATCAAGAAAGAGACAGCAACCAAGTGTGCGCAAGCTGTGTGAC	900
D	b	841	AAACAACCCCGCTGTTCGCGCATCAAGAAAGAGACAGCAACCAAGTGTGCGCAAGCTGTGTGAC	900
Q	y	901	TTCCGCGAGCTGAACAAGGCGACCCAGGACTTGTGAGAGGTGACGTGTGGGCAATCCCCCAG	960
D	b	901	TTCCGCGAGCTGAACAAGGCGACCCAGGACTTGTGAGAGGTGACGTGTGGGCAATCCCCCAG	960
Q	y	961	CCCGCGCGCTGAAGAAAGAAAGACGCTGACGCTGTGGAAGTGTGGGCGACGCTACTTC	1020
D	b	961	CCCGCGCGCTGAAGAAAGAAAGACGCTGACGCTGTGGAAGTGTGGGCGACGCTACTTC	1020
Q	y	1021	AGCGTGCCTCTGAGCGAGGACTTTCGCGAAGTACACCGGCTTTCACATCCCAAGCATCAAC	1080
D	b	1021	AGCGTGCCTCTGAGCGAGGACTTTCGCGAAGTACACCGGCTTTCACATCCCAAGCATCAAC	1080
Q	y	1081	AAACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTTGAAAGGGCAGC	1140
D	b	1081	AAACGAGACCCCGCGCATCCGCTACCAAGTACAAAGTGTGCTGCCCCAGGCTTGAAAGGGCAGC	1140
Q	y	1141	CCCAAGCATCTTCCAGAGCAGACGACCAAGATCTGTGAAGCCMTTCGCGCGCGCGCAACCCC	1200
D	b	1141	CCCAAGCATCTTCCAGAGCAGACGACCAAGATCTGTGAAGCCMTTCGCGCGCGCGCAACCCC	1200
Q	y	1201	GAGATCGTATCTACCAAGTACATGAGACGACTGTGAGTGTGGCGAGCATCTGAGAGATGTGC	1260
D	b	1201	GAGATCGTATCTACCAAGTACATGAGACGACTGTGAGTGTGGCGAGCATCTGAGAGATGTGC	1260
Q	y	1261	CAGCACCGCGCCAGATCGAGAGCTGTGCGACAGCAACCTGTGTGTGTGGGCTTTCACCAAC	1320

Db	1261	CAGCAACGGCGCAAGATCGAGAGACTGCGCAGACACTGCTGGCTGGGGCTTCAACACC	1320
Qy	1321	CCCGACAAGAGACCGAAGAGAGCCCCCTTCTCTGTGATGATGCTACAGACTGCACCC	1380
Db	1321	CCCGACAAGAGACCGAAGAGAGCCCCCTTCTCTGTGATGATGCTACAGACTGCACCC	1380
Qy	1381	GACAAATGAGACCGGTGACGCCCATGACACTGCCGAGAAAGAGAGCTGACACCTGTAAAGAC	1440
Db	1381	GACAAATGAGACCGGTGACGCCCATGACACTGCCGAGAAAGAGAGCTGACACCTGTAAAGAC	1440
Qy	1441	ATCCAGAAGCTGTGTGGGCAAGCTGAATCTGTGGGCGACGAGACTTACCCCCGGCATCAAGTG	1500
Db	1441	ATCCAGAAGCTGTGTGGGCAAGCTGAATCTGTGGGCGACGAGACTTACCCCCGGCATCAAGTG	1500
Qy	1501	CGCCAGCTGTGCAAGCTGTGTGGGCGGCGCAAGGCCCTTGACCGAGCATGTGTCCCTGACC	1560
Db	1501	CGCCAGCTGTGCAAGCTGTGTGGGCGGCGCAAGGCCCTTGACCGAGCATGTGTCCCTGACC	1560
Qy	1561	GAGAGAGCCCGAGCTGTGAGCTGTGCCGAGAACCGCGAGATCTCTGTGGCGAGCCCTGTGACGGC	1620
Db	1561	GAGAGAGCCCGAGCTGTGAGCTGTGCCGAGAACCGCGAGATCTCTGTGGCGAGCCCTGTGACGGC	1620
Qy	1621	GTGTACTACGACCCCGACGAAAGGACCTGTGTGGCGAGATCCGAGAAAGGAGGCGCACGACAG	1680
Db	1621	GTGTACTACGACCCCGACGAAAGGACCTGTGTGGCGAGATCCGAGAAAGGAGGCGCACGACAG	1680
Qy	1681	TGGACTTACCAAGATCTACCAAGAGCCCTTCAAGAACCTTGAAAGCCGACCAAGTACGGCAAG	1740
Db	1681	TGGACTTACCAAGATCTACCAAGAGCCCTTCAAGAACCTTGAAAGCCGACCAAGTACGGCAAG	1740
Qy	1741	ATGTGCAACCGGCCCAACCAACAGATGTGAAGCACTGACCCGAGGCCGTGTACAGAAAGTGGCC	1800
Db	1741	ATGTGCAACCGGCCCAACCAACAGATGTGAAGCACTGACCCGAGGCCGTGTACAGAAAGTGGCC	1800
Qy	1801	ATGTGAGAGCATCGAGATCTGTGGGCAAGACCCCGAAGTTCCGCTGGCCATCCAGAAAGAG	1860
Db	1801	ATGTGAGAGCATCGAGATCTGTGGGCAAGACCCCGAAGTTCCGCTGGCCATCCAGAAAGAG	1860
Qy	1861	ACCTGTGAGAGCACTGTGTGAGCCGACCTACTGTGAGAGGCCACCTTGATCCCCGAGTGTGAATTCC	1920
Db	1861	ACCTGTGAGAGCACTGTGTGAGCCGACCTACTGTGAGAGGCCACCTTGATCCCCGAGTGTGAATTCC	1920
Qy	1921	GTGAACACCCCCCTGTGTGAGACTGTGTGTACAGCTGTGAGAAAGAGGCCCATATGTGGC	1980
Db	1921	GTGAACACCCCCCTGTGTGAGACTGTGTGTACAGCTGTGAGAAAGAGGCCCATATGTGGC	1980
Qy	1981	GGCGAGACCTTCTACGTGTGAACGGGCGCGCCAACCGCGAGACCAAGATCCGCAAGGCGGCGC	2040
Db	1981	GGCGAGACCTTCTACGTGTGAACGGGCGCGCCAACCGCGAGACCAAGATCCGCGAAGGCGGCGC	2040
Qy	2041	TACGTGACCGACCGGGGCGCGCAGAAAGATCGTGAAGCTGACCGAGACCAACCAACAGAG	2100
Db	2041	TACGTGACCGACCGGGGCGCGCAGAAAGATCGTGAAGCTGACCGAGACCAACCAACAGAG	2100
Qy	2101	ACCGAGCTGTGACGATTCAGCTGTGGCCCTTGACGAGACAGCGGACGAGAGTGAACATGTGTG	2160
Db	2101	ACCGAGCTGTGACGATTCAGCTGTGGCCCTTGACGAGACAGCGGACGAGAGTGAACATGTGTG	2160
Qy	2161	ACCGACAGCCAGTACGGCTGTGGAGCATATCCAGGGCCGAGCCCGACAAGAGGAGAGGAGG	2220
Db	2161	ACCGACAGCCAGTACGGCTGTGGAGCATATCCAGGGCCGAGCCCGACAAGAGGAGAGGAGG	2220
Qy	2221	CTGTGTGAACAGATCATTCAGACACTGTATCAAGAGAGAGAGAGGTGTACTGTAGCTGTGGGTG	2280
Db	2221	CTGTGTGAACAGATCATTCAGACACTGTATCAAGAGAGAGAGAGGTGTACTGTAGCTGTGGGTG	2280
Qy	2281	CCCGCCCAAGAGGCGATTCGGCGGCAACGAGCAGATCGACAAGCTGTGTAGCGAAGGCGATC	2340
Db	2281	CCCGCCCAAGAGGCGATTCGGCGGCAACGAGCAGATCGACAAGCTGTGTAGCGAAGGCGATC	2340
Qy	2341	CGCAAGGTGCTGTTCTGTGACGGCATGTATGGCGGACTGTGATCTACAGTACATGTGAC	2400
Db	2341	CGCAAGGTGCTGTTCTGTGACGGCATGTATGGCGGACTGTGATCTACAGTACATGTGAC	2400

QY 2401 GACCTGTAGTGGGAGCGCGCGCCCTTAGAGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
DB 2401 GACCTGTAGTGGGAGCGCGCGCCCTTAGAGATCGATTAAAGCTTCCCGGGCTAGCAC 2460
QY 2461 GGTAATTTC 2469
DB 2461 GGTAATTTC 2469

RESULT 2
AX455915
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.

REFERENCE
1 zur Megeide, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
AUTHORS Polynucleotides encoding antigenic hiv type c polypeptides,
TITLE polypeptides and uses thereof
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source 1..2463
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN
Query Match 98.9%; Score 2442.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 1.4e-201;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GTGACGCCACCATGCGCCGACATGAGCCAGGCCACGAGGCCAACAATCTTGATGAG 60
DB 1 GTGACGCCACCATGCGCCGACATGAGCCAGGCCACGAGGCCAACAATCTTGATGAG 60
QY 61 CGCAGCAATCTCAAGGGGCCCCCAAGCCGATCATTAAGTGTCTTCAACTGCGCGCAAGAGGGC 120
DB 61 CGCAGCAATCTCAAGGGGCCCCCAAGCCGATCATTAAGTGTCTTCAACTGCGCGCAAGAGGGC 120
QY 121 CACATGCGCGCAACTGCGCGCGCGCCGCAAGAGGGCTGTGAAAGTGGCGCGCAAGAG 180
DB 121 CACATGCGCGCAACTGCGCGCGCGCCGCAAGAGGGCTGTGAAAGTGGCGCGCAAGAG 180
QY 181 GGCACACAGATGAGGACTGCAACGAGCGCCAGGCCAATCTTCCGCGAGGACTTGAGC 240
DB 181 GGCACACAGATGAGGACTGCAACGAGCGCCAGGCCAATCTTCCGCGAGGACTTGAGC 240
QY 241 TTCCCCCAAGGCAAGGCGCGCGAGTTCCCAAGGAGAGAACCGGCGCAACGCCCCACC 300
DB 241 TTCCCCCAAGGCAAGGCGCGCGAGTTCCCAAGGAGAGAACCGGCGCAACGCCCCACC 300
QY 241 TTCCCCCAAGGCAAGGCGCGCGAGTTCCCAAGGAGAGAACCGGCGCAACGCCCCACC 300
DB 241 TTCCCCCAAGGCAAGGCGCGCGAGTTCCCAAGGAGAGAACCGGCGCAACGCCCCACC 300
QY 301 AGCGCGAGCTGAGGTGCGCGCGCAACCCCGGAGGAGCGCGCGCGCGCGCGCGAG 360
DB 301 AGCGCGAGCTGAGGTGCGCGCGCAACCCCGGAGGAGCGCGCGCGCGCGCGCGAG 360
QY 361 GGCACCTTGAATCTTCCCGAGATCACTGTGTGCAAGCGCCCTGTGTGAGCATCAAGTG 420
DB 361 GGCACCTTGAATCTTCCCGAGATCACTGTGTGCAAGCGCCCTGTGTGAGCATCAAGTG 420
QY 421 GCGCGCAAGTCAAGAGGCGCTGTGTGAGCAACCGCGCGCGCGCAACCGGTGTGAGAG 480
DB 421 GCGCGCAAGTCAAGAGGCGCTGTGTGAGCAACCGCGCGCGCGCAACCGGTGTGAGAG 480
QY 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATGCGGGGCTTCATCAAG 540
DB 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATGCGGGGCTTCATCAAG 540

QY 541 GTGCGCGAGTACGACCAAGATCTTGATTCGAGATCTGCGCGAGAGAGGCCATCGGCAACCG 600
DB 541 GTGCGCGAGTACGACCAAGATCTTGATTCGAGATCTGCGCGAGAGAGGCCATCGGCAACCG 600
QY 601 CTGATCGGCCCCCAGCCCGGTGAACATCAATCGGCGCGCAACATCTGAGCCAGGCTGGCTGC 660
DB 601 CTGATCGGCCCCCAGCCCGGTGAACATCAATCGGCGCGCAACATCTGAGCTGGCTGC 660
QY 661 ACCCTGAATCTTCCCATGAGCCCATGAGACCGTGCCTGTAAGCTGAAGCCCGGCAATG 720
DB 661 ACCCTGAATCTTCCCATGAGCCCATGAGACCGTGCCTGTAAGCTGAAGCCCGGCAATG 720
QY 721 GACGCGCCCCAGAGTGAAGAGTGGCGCTGTGACCCGAGAGAGAAATCAAGGCCCTGACCGC 780
DB 721 GACGCGCCCCAGAGTGAAGAGTGGCGCTGTGACCCGAGAGAGAAATCAAGGCCCTGACCGC 780
QY 781 ATCTGAGAGAGATGAGAGAGAGGCAAGATCAACCAAGATGCGGCCCGAGAACCCCTAC 840
DB 781 ATCTGAGAGAGATGAGAGAGAGGCAAGATCAACCAAGATGCGGCCCGAGAACCCCTAC 840
QY 841 AACACCCCGGTGTTCGCTCAAGAGAGAGAGCAAGCAACCAAGTGGCGCAAGCTGTGAGAC 900
DB 841 AACACCCCGGTGTTCGCTCAAGAGAGAGAGCAAGCAACCAAGTGGCGCAAGCTGTGAGAC 900
QY 901 TTCCCGAGCTGAACAAGCGCAACCGAGCTTCTGTGAGAGTGCAGCTGGGCTATCCCGCAC 960
DB 901 TTCCCGAGCTGAACAAGCGCAACCGAGCTTCTGTGAGAGTGCAGCTGGGCTATCCCGCAC 960
QY 961 CCGCGCGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGAGAGTGGCGAGCCGCTACTTC 1020
DB 961 CCGCGCGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGAGAGTGGCGAGCCGCTACTTC 1020
QY 1021 AGCGTGCCTGTGAAGAGAGCTTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080
DB 1021 AGCGTGCCTGTGAAGAGAGCTTTCGCAAGTACACCGCTTCAACATCCCGAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCAATCCGCTACCAAGTACAGTGTCTGCGCGAGGCTGAAAGGGCAGC 1140
DB 1081 AACGAGACCCCGGCAATCCGCTACCAAGTACAGTGTCTGCGCGAGGCTGAAAGGGCAGC 1140
QY 1141 CCAGAGCTTTCAGAGCGAGTGAACCAAGATCTGAGAGCCCTTCGCGCGCGCAACCC 1200
DB 1141 CCAGAGCTTTCAGAGCGAGTGAACCAAGATCTGAGAGCCCTTCGCGCGCGCAACCC 1200
QY 1201 GAGATGTGATCTACCAAGTACATGAGCAAGCTGTGAGTGGCGAGCACTGTGAGATCGCG 1260
DB 1201 GAGATGTGATCTACCA-----GCGCGCGCTGTGAGTGGCGAGCACTGTGAGATCGCG 1260
QY 1261 CAGCACCGCGCAAGATCGAGAGAGCTGCGCAAGCACCTGCTGCGCTTTCACACCC 1320
DB 1261 CAGCACCGCGCAAGATCGAGAGAGCTGCGCAAGCACCTGCTGCGCTTTCACACCC 1320
QY 1321 CCGGCAAGAACCAACCAAGAGAGCGCCCTTCTGTGATGAGGCTGACAGAGCTGCAACCC 1380
DB 1321 CCGGCAAGAACCAACCAAGAGAGCGCCCTTCTGTGATGAGGCTGACAGAGCTGCAACCC 1380
QY 1315 CCGGCAAGAACCAACCAAGAGAGCGCCCTTCTGTGATGAGGCTGACAGAGCTGCAACCC 1374
DB 1315 CCGGCAAGAACCAACCAAGAGAGCGCCCTTCTGTGATGAGGCTGACAGAGCTGCAACCC 1374
QY 1381 GACAAATGAGACCTGTGAGCGCCATTCGAGCTGCGCGAGAGAGAGAGTGTGACCCGTAACGAC 1440
DB 1381 GACAAATGAGACCTGTGAGCGCCATTCGAGCTGCGCGAGAGAGAGAGTGTGACCCGTAACGAC 1440
QY 1375 GACAAATGAGACCTGTGAGCGCCATTCGAGCTGCGCGAGAGAGAGTGTGACCCGTAACGAC 1434
DB 1375 GACAAATGAGACCTGTGAGCGCCATTCGAGCTGCGCGAGAGAGAGTGTGACCCGTAACGAC 1434
QY 1441 ATTCAGAAAGCTGTGAGCAAGCTGAACCTGAGCGCAAGCAGATCTACCCGCGATCAAGTG 1500
DB 1441 ATTCAGAAAGCTGTGAGCAAGCTGAACCTGAGCGCAAGCAGATCTACCCGCGATCAAGTG 1500
QY 1435 ATTCAGAAAGCTGTGAGCAAGCTGAACCTGAGCGCAAGCAGATCTACCCGCGATCAAGTG 1494
DB 1435 ATTCAGAAAGCTGTGAGCAAGCTGAACCTGAGCGCGCAAGCAGATCTACCCGCGATCAAGTG 1494
QY 1501 CCGCGAGCTGTGAAGTGTGTGCGCGCGCGCAAGCGCTGACCAAGTGTGAGCGCGCGAGC 1560
DB 1501 CCGCGAGCTGTGAAGTGTGTGCGCGCGCGCAAGCGCTGACCAAGTGTGAGCGCGCGAGC 1560
QY 1495 CCGCGAGCTGTGAAGTGTGTGCGCGCGCGCAAGCGCTGACCAAGTGTGAGCGCGCGAGC 1554
DB 1495 CCGCGAGCTGTGAAGTGTGTGCGCGCGCGCAAGCGCTGACCAAGTGTGAGCGCGCGAGC 1554
QY 1561 GAGAGAGCGAGCTGTGAGCTGCGCGAGAACCCGAGATCTGCGCGAGCGCGCTGCAAGC 1620
DB 1561 GAGAGAGCGAGCTGTGAGCTGCGCGAGAACCCGAGATCTGCGCGAGCGCGCTGCAAGC 1620
QY 1555 GAGAGAGCGAGCTGTGAGCTGCGCGAGAACCCGAGATCTGCGCGAGCGCGCTGCAAGC 1614
DB 1555 GAGAGAGCGAGCTGTGAGCTGCGCGAGAACCCGAGATCTGCGCGAGCGCGCTGCAAGC 1614

QY 1621 GTGTACTAGACCCGAGCAGAGGACTGTGGCCGAGATCCAGAGCAGGCGCAGCAG 1680
 Db 1615 GTGTACTAGACCCGAGCAGAGGACTGTGGCCGAGATCCAGAGCAGGCGCAGCAG 1674
 QY 1661 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAAGAATCCGCGCAAGTACCCCAAG 1740
 Db 1675 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAAGAATCCGCGCAAGTACCCCAAG 1734
 QY 1741 ATGCGCAGCGCCCAACCAACGAGCTGAAAGCAGTGAACGAGAGCCGTGAGAGAAATCCGC 1800
 Db 1735 ATGCGCAGCGCCCAACCAACGAGCTGAAAGCAGTGAACGAGAGCCGTGAGAGAAATCCGC 1794
 QY 1801 ATGAGAGCATGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGAG 1860
 Db 1795 ATGAGAGCATGTGATCTGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGAG 1854
 QY 1861 ACCTGGAGAGCCTGGTGGACCGACTACTAGAGGCGCACTTGATATCCCGAGTGGAGTTTC 1920
 Db 1855 ACCTGGAGAGCCTGGTGGACCGACTACTAGAGGCGCACTTGATATCCCGAGTGGAGTTTC 1914
 QY 1921 GTGAACACCCCGCTGGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCCATCCGC 1980
 Db 1915 GTGAACACCCCGCTGGTGAAGCTGTGTGTACAGCTGAGAGAGAGCCCATCCATCCGC 1974
 QY 1981 GCGAGACCTTCTACGTGAGACGCGCGCCCAACCGGAGACCAAGATCCGCAAGGCCCGGC 2040
 Db 1975 GCGAGACCTTCTACGTGAGACGCGCGCCCAACCGGAGACCAAGATCCGCAAGGCCCGGC 2034
 QY 2041 TACGTGACCGACCGGGGCGCGCAAGATCTGTGAGCTTGAACCCAGACCAACCAAG 2100
 Db 2035 TACGTGACCGACCGGGGCGCGCAAGATCTGTGAGCTTGAACCCAGACCAACCAAG 2094
 QY 2101 ACCGAGCTGACAGGCGCATCCAGCTGGCCCTGACAGACAGCGGCGAGCGAGTGAACATCCGTG 2160
 Db 2095 ACCGAGCTGACAGGCGCATCCAGCTGGCCCTGACAGACAGCGGCGAGCGAGTGAACATCCGTG 2154
 QY 2161 ACCGACAGCAGTACGCGCTTGCGCATCATCAGGCCAGCCCGCAACAGACGAGACGAG 2220
 Db 2155 ACCGACAGCAGTACGCGCTTGCGCATCATCAGGCCAGCCCGCAACAGACGAGACGAG 2214
 QY 2221 CTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGGTG 2280
 Db 2215 CTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGGTG 2274
 QY 2281 CCGCGCCCAAGAGGCGATCGGCGGCAACGAGCATCGCAAGCTGGTGAAGAGGCGCATC 2340
 Db 2275 CCGCGCCCAAGAGGCGATCGGCGGCAACGAGCATCGCAAGCTGGTGAAGAGGCGCATC 2334
 QY 2341 CGCAAGGTGCTGTTCTTGACGCGCATCGATGCGCGCATCTGATCTACCAAGTACATGAGC 2400
 Db 2335 CGCAAGGTGCTGTTCTTGACGCGCATCGATGCGCGCATCTGATCTACCAAGTACATGAGC 2394
 QY 2401 GACCTGTACGTGGGCGAGCGGCGCCCTTAGATCGATTAAAGCTTCCGCGGCTGACGCC 2460
 Db 2395 GACCTGTACGTGGGCGAGCGGCGCCCTTAGATCGATTAAAGCTTCCGCGGCTGACGCC 2454
 QY 2461 GGTGAATTC 2469
 Db 2455 GGTGAATTC 2463

RESULT 3
 AX455916 2457 bp DNA linear PAT 06-JUL-2002
 LOCUS AX455916
 DEFINITION Sequence 32 from Patent WO0204493.
 ACCESSION AX455916
 VERSION AX455916.1 GI:21714901
 KEYWORDS
 ORGANISM
 SOURCE
 synthetic construct
 other sequences; artificial sequences.
 REFERENCE
 1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
 JOURNAL polyptides and uses thereof
 CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
 FEATURES
 source
 1. 2457
 Location/Qualifiers
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="PR975YMMM"
 ORIGIN
 Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
 Best Local Similarity 99.3%; Pred. No. 2.7e-199;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
 QY 1 GTGACGCGCACCATGTGGCGAGGCGCATGTAGCCAGGCGCACAGGCGCAATCTGTATGCG 60
 Db 1 GTGACGCGCACCATGTGGCGAGGCGCATGTAGGCTTCACTGCGGCAAGAGGCG 60
 QY 61 CGCAGCAACTTCAAGAGGCGCCCAAGGCGCATCATAGTGCTTCAACTGCGGCAAGAGGCG 120
 Db 61 CGCAGCAACTTCAAGAGGCGCCCAAGGCGCATCATAGTGCTTCAACTGCGGCAAGAGGCG 120
 QY 121 CACATGCGCCCGCACTGCGCGCCCGCCGCAAGAGGCGTGTGAAAGTGCAGCAAGAG 180
 Db 121 CACATGCGCCCGCACTGCGCGCGCCCGCCGCAAGAGGCGTGTGAAAGTGCAGCAAGAG 180
 QY 181 GCGCACGATGAGAGACTGTGACCGAGCGCCCAACTTTCTTCCGGAAGACTTGCGCC 240
 Db 181 GCGCACGATGAGAGAGACTGTGACCGAGCGCCCAACTTTCTTCCGGAAGAGACTTGCGCC 240
 QY 241 TTCCCCCGAGGCAAGGCGCGCGAGTTCCCGAGCGAGCAAGACCGGCGCAACGCGCCACC 300
 Db 241 TTCCCCCGAGGCAAGGCGCGCGAGTTCCCGAGCGAGCAAGACCGGCGCAACGCGCCACC 300
 QY 301 AGCCGAGCTGCAAGTGCAGCGCAACACCCCGCAGCGAGCGCGCGCGAGCGCCAG 360
 Db 301 AGCCGAGCTGCAAGTGCAGCGCAACACCCCGCAGCGAGCGCGCGCGAGCGCGCGAG 360
 QY 361 GGCACCTGGAATTTCCCGCAATCAACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGTG 420
 Db 361 GGCACCTGGAATTTCCCGCAATCAACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTGTG 420
 QY 421 GCGCGCCAGATCAAGAGGCGCTGTGCAACCGGCGCGCAACCGGTGTTGAGAGAG 480
 Db 421 GCGCGCCAGATCAAGAGGCGCTGTGCAACCGGCGCGCAACCGGTGTTGAGAGAGAG 480
 QY 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGCATCGCGCTTCAATCAAG 540
 Db 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCGGCGCATCGGCGCTTCAATCAAG 540
 QY 541 GTGCGCCAGTACCAAGATCTGTATCGAGATCTGCGGCAAGAGAGGCGCATCGGCGCGT 600
 Db 541 GTGCGCCAGTACCAAGATCTGTATCGAGATCTGCGGCAAGAGAGGCGCATCGGCGCGT 600
 QY 601 CTGATGCGCCCGCACCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
 Db 601 CTGATGCGCCCGCACCCCGGTGAACATCATCGGCGCGCAACATGCTGACCCAGCTGGGCTGC 660
 QY 661 ACCCTGAATCTCCCATGAGCCCATCGAGACCGGTGCGGAGCTGAAGCCCGGCAATG 720
 Db 661 ACCCTGAATCTCCCATGAGCCCATCGAGACCGGTGCGGAGCTGAAGCCCGGCAATG 720
 QY 721 GACGCGCCCAAGTGAAGAGTGTGCGCCCTGACCGAGAGAGAGATCAAGGCGCTTGAACCGGC 780
 Db 721 GACGCGCCCAAGTGAAGAGTGTGCGCCCTGACCGAGAGAGAGATCAAGGCGCTTGAACCGGC 780
 QY 781 ATTCGCGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCGCGAGAACCCCTAC 840
 Db 781 ATTCGCGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCGCGAGAACCCCTAC 840
 QY 841 AACACCCCGTGTGCGCATCAAGAGAGAGACCAACAGTGGCGCAAGCTGTGTGAGC 900


```

Db      841 AACACCCCGTGTGGCCATCAAGAAAGAGACCAAGTGGCGCAAGCTGTGAGC 900
Qy      901 TTCCCGAGCTGAACAAGGCAACCAAGACTTCTGGAGGTGCACTGGGATCCCCAC 960
Db      901 TTCCCGAGCTGAACAAGGCAACCAAGACTTCTGGAGGTGCACTGGGATCCCCAC 960
Qy      961 CCCCGCGCTTGAAAGAAAGAGAGCTGACCTGTGTGAGAGTGGGGCGACCTTACTTC 1020
Db      961 CCCCGCGCTTGAAAGAAAGAGAGCTGACCTGTGTGAGAGTGGGGCGACCTTACTTC 1020
Qy      1021 AGCGTGGCCCTTGAGAGAGACTTCCGCAAGTACACCGCTTCACCATCCCGACATCAAC 1080
Db      1021 AGCGTGGCCCTTGAGAGAGACTTCCGCAAGTACACCGCTTCACCATCCCGACATCAAC 1080
Qy      1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAACTGTGCCCCAGGGCTGAAAGGGCAGC 1140
Db      1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAACTGTGCCCCAGGGCTGAAAGGGCAGC 1140
Qy      1141 CCCAGCATTTTCAAGAGACGATGACCAAGATCTGAGAGCTTCCGCGCCGCAACCC 1200
Db      1141 CCCAGCATTTTCAAGAGACGATGACCAAGATCTGAGAGCTTCCGCGCCGCAACCC 1200
Qy      1201 GAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGAGCGACTTGAGATCGGC 1260
Db      1201 GAGATCTGTATCTACCA-----GGCCCCCTGTACGTGGGAGCGACTTGAGATCGGC 1260
Qy      1261 CAGCACCGCGCAAGATCGAGAGAGTGGCGCAAGACTGTGCGTGGGGCTTACCAACC 1320
Db      1255 CAGCACCGCGCAAGATCGAGAGAGTGGCGCAAGACTGTGCGTGGGGCTTACCAACC 1314
Qy      1321 CCCGACAGAGACACAGAAAGAGCCCTTCTGTGTGAGTGGGCTACAGCTGCACCCC 1380
Db      1315 CCCGACAGAGACACAGAAAGAGCCCTTCTGTGTGAGTGGGCTACAGCTGCACCCC 1368
Qy      1381 GACAGTGGAGCCGTGAGCCCATTCGAGCTGCCCGAGAGAGAGAGTGGACCGGTAAGAC 1440
Db      1369 GACAGTGGAGCCGTGAGCCCATTCGAGCTGCCCGAGAGAGAGAGTGGACCGGTAAGAC 1428
Qy      1441 ATTCAGAAAGCTGTGGGCAAGCTGAACTGGGCAAGCCAGATCTACCCCGCATCAAGGTG 1500
Db      1429 ATTCAGAAAGCTGTGGGCAAGCTGAACTGGGCAAGCCAGATCTACCCCGCATCAAGGTG 1488
Qy      1501 CGCCAGCTGTGCAAGCTGTGCGCGGCGCAAGAGCCCTGACCGCATGTGCCCCCTGACC 1560
Db      1489 CGCCAGCTGTGCAAGCTGTGCGCGGCGCAAGAGCCCTGACCGCATGTGCCCCCTGACC 1548
Qy      1561 GAGAGAGCCGAGCTGAGAGCTGGCCGAGAACCGGAGATCTGTGCGAGAGCCCTGTGCAAGC 1620
Db      1549 GAGAGAGCCGAGCTGAGAGCTGGCCGAGAACCGGAGATCTGTGCGAGAGCCCTGTGCAAGC 1608
Qy      1621 GTGTACTAGACCCCGCAAGAGACTGTGGCCGAGATCCAGAAAGAGGGCCACAGCACAG 1680
Db      1609 GTGTACTAGACCCCGCAAGAGACTGTGGCCGAGATCCAGAAAGAGGGCCACAGCACAG 1668
Qy      1681 TGGACTTACAGATCTACCAAGAGAGCCCTTCAAGAACCTGAAAGCCGGCAAGTACGCCAAG 1740
Db      1669 TGGACTTACAGATCTACCAAGAGAGCCCTTCAAGAACCTGAAAGCCGGCAAGTACGCCAAG 1728
Qy      1741 ATGGCGACCGGCGCAACCAAGAGCTGAAAGCACTGACCGAGGCGCTGCAAGAAATCGCC 1800
Db      1729 ATGGCGACCGGCGCAACCAAGAGCTGAAAGCACTGACCGAGGCGCTGCAAGAAATCGCC 1788
Qy      1801 ATGAGAGAGATCTGTATCTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db      1789 ATGAGAGAGATCTGTATCTGGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848
Qy      1861 ACCTGGAAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db      1849 ACCTGGAAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1908
Qy      1921 GTGAAACACCCCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980

```

```

Db      1909 GTGAAACACCCCGCTGTGTGAGAGCTGTGTATCAAGCTGAGAAAGAGAGCCATCGGC 1968
Qy      1981 GCCGAGACCTTCTAGTGAAGGCGCGCCCAACCGCGAGACCAAGATGGCAAGGCCCGGC 2040
Db      1969 GCCGAGACCTTCTAGTGAAGGCGCGCCCAACCGCGAGACCAAGATGGCAAGGCCCGGC 2028
Qy      2041 TACGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db      2029 TACGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
Qy      2101 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db      2089 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2148
Qy      2161 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db      2149 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2208
Qy      2221 CTGTGAAACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db      2209 CTGTGAAACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2268
Qy      2281 CCCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db      2269 CCCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2328
Qy      2341 CGCAAGGTGTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db      2329 CGCAAGGTGTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2388
Qy      2401 GACCTGTACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Db      2389 GACCTGTACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2448
Qy      2461 GGTGAATTC 2469
Db      2449 GGTGAATTC 2457

RESULT 4
BD263706          2312 bp      DNA      linear      PART 17-JUL-2003
LOCUS             BD263706
DEFINITION       Improved expression of HIV polypeptides and production of
                  virus-like particles.
ACCESSION        BD263706
VERSION          BD263706.1 GI:33073474
KEYWORDS         JP 2002533124-A/73.
SOURCE           synthetic construct
ORGANISM         other sequences; artificial sequences.
REFERENCE        1 (bases 1 to 2312)
AUTHORS          Barnett,S., Megede,J.Z., Sriwasatava,I., Lian,Y., Hartog,K., Liu,H.,
                  Greer,C., Selby,M. and Walker,C.
TITLE            Improved expression of HIV polypeptides and production of
                  virus-like particles
JOURNAL          Patent: JP 2002533124-A 73 08-OCT-2002;
                  CHIRON CORP
COMMENT          OS Artificial Sequence
                  PN JP 2002533124-A/73
                  PD 08-OCT-2002
                  PP 30-DEC-1999 JP 2000591193
                  PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
                  SUSAN BARNETT,JAN ZUR MEGEDE,INDRISH SRIWASATAVA,YING LIAN, PI
                  KAREN HARTOG,
                  PI HONG LIU,CATHERINE GREER,MARK SELBY,CHRISTOPHER WALKER PC
                  C12N15/09,A61K31/711,A61K48/00,A61P31/18,A61P37/02, PC
                  C12N5/10,
                  PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
                  Description of Artificial Sequence: FS(-).protomod.Rtopr(+) FH
                  Location/Qualifiers
                  FT source 1..2312
                  Location/Qualifiers
                  FEATURES

```

Source 1. 2312
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

Query Match 83.1%; Score 2052; DB 6; Length 2312;
Best Local Similarity 93.6%; Pred. No. 4,9e-168;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

ORIGIN
170 GCGGCAAGAGGGCCCAAGATGAGAGTGCACCGAGGCGCAAGCTTCTCCGCG 229
Db 1 GCGGCGCGAAGAGCAACCAATGAAAGATTGCACTGAGAGACAGGCTTAATTTCTCCGCG 60
Qy 230 AGGACTGTGGCTTCTCCGAGGGCAAGGCGCGAGTTCCCGAGCGAGCAAGACCGGCGCA 289
Db 61 AGGACTGTGGCTTCTCCGAGGGCAAGGCGCGAGTTCCCGAGCGAGCAAGACCGGCGCA 120
Qy 290 AAGGCCCCCAAGCGCGAGCTGCAAGGTGCGGCGG-----ACAACCCCGCAGCGAGG 343
Db 121 AAGGCCCCCAAGCGCGAGCTGCAAGGTGCGGCGGCGAGCAAGACCGCTGAGCGAGG 180
Qy 344 CCGGCGCGAGGCGCGAGGCAACCTG-----AATTTCCCGCAATCACTCTGTGGCAGC 397
Db 181 CCGGCGCGAGGCGCGAGGCAACCTG-----AATTTCCCGCAATCACTCTGTGGCAGC 240
Qy 398 GCGGCGCGAGGCAATCAAGGTGGGCGGCGAGATCAAGAGGCGCTGTGCAACCGGCG 457
Db 241 GCGGCGCGAGGCAATCAAGGTGGGCGGCGAGATCAAGAGGCGCTGTGCAACCGGCG 300
Qy 458 CCGAGCAACCGGTGTGAGAGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCG 517
Db 301 CCGAGCAACCGGTGTGAGAGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGATCG 360
Qy 518 GCGGCAATGCGCGCTTTCATCAAGGTGCGCGCAATGAGCAAGTCTTGAATCTGCG 577
Db 361 GCGGCAATGCGCGCTTTCATCAAGGTGCGCGCAATGAGCAAGTCTTGAATCTGCG 420
Qy 578 GCAAGAGGCGCATCGGCAACGTGTGATCGGCGCGCAACCGCGTGAACATCATCGGCGCA 637
Db 421 GCGGCAAGGCGCATCGGCAACGTGTGATCGGCGCGCAACCGCGTGAACATCATCGGCGCA 480
Qy 638 AATGCTGAGCCAGGTGGCTGCAACCTGGAATTTCCCGCATGAGCCCATGAGACCGTGC 697
Db 481 AATGCTGAGCCAGGTGGCTGCAACCTGGAATTTCCCGCATGAGCCCATGAGACCGTGC 540
Qy 698 CCGTGAAGCTGAAGCCCGGCAATGAGAGGCGCCCAAGTGAAGAGTGGCCCTTGAACGAG 757
Db 541 CCGTGAAGCTGAAGCCCGGCAATGAGAGGCGCCCAAGTGAAGAGTGGCCCTTGAACGAG 600
Qy 758 AGAAGATCAAGGCGCTGACCGGCATCTGCGAGAGATGAGAGAGGCGCAAGATCAACA 817
Db 601 AGAAGATCAAGGCGCTGCGTGAAGATCTGACCGAGTGAAGAGAGGCGCAAGATCAACA 660
Qy 818 AGATCGGCGCGCGAGAACCCCTTCAACAACCCCGTGTTCGCAATCAAGAGAGAGCAGCA 877
Db 661 AGATCGGCGCGCGAGAACCCCTTCAACAACCCCGTGTTCGCAATCAAGAGAGAGCAGCA 720
Qy 878 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGAGTGAACAAGCGCACCCAGACTTCTGCG 937
Db 721 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGAGTGAACAAGCGCACCCAGACTTCTGCG 780
Qy 938 AGGTGAGCTGAGGATCCCGCAACCGCGCGGCGCTGAAGAGAGAGAGACGTGACCGTGC 997
Db 781 AGGTGAGCTGAGGATCCCGCAACCGCGCGGCGCTGAAGAGAGAGAGACGTGACCGTGC 840
Qy 998 TGAAGTGGCGCAAGCTGTGAGCTTCAAGCGTGCCTTGAAGAGAGCTTCCGCAAGTCAACCG 1057
Db 841 TGAAGTGGCGCAAGCTGTGAGCTTCAAGCGTGCCTTGAAGAGAGCTTCCGCAAGTCAACCG 900
Qy 1058 CTTTCAACCATCCCGAGATGAACAAGAGACCCCGGCGCATTCGCTTCAAGTCAACAGTGC 1117
Db 901 CTTTCAACCATCCCGAGATGAACAAGAGACCCCGGCGCATTCGCTTCAAGTCAACAGTGC 960

Qy 1118 TGCCCCAGAGGCTGAAAGGCGAGCCCGAGCATCTTCCAGAGCAGATGACCAAGATCTTGG 1177
Db 961 TGCCCCAGAGGCTGAAAGGCGAGCCCGGCGCATCTTCCAGAGCAGATGACCAAGATCTTGG 1020
Qy 1178 AGGCTTTCGCGCGCGCAACCCCGAGATCTGATCTCAAGTCAATGAGCAAGCTGTACG 1237
Db 1021 AGGCTTTCGCGAGAGAACCCCGCATCTGTGATCTCAAGTCAATGAGCAAGCTGTACG 1080
Qy 1238 TGAGGAGGACCTGAGATTCGGCGAGCAACCGGCGCAAGATGAGAGAGCTTGGCGAAGCC 1297
Db 1081 TGAGGAGGACCTGAGATTCGGCGAGCAACCGGCGCAAGATGAGAGAGCTTGGCGAAGCC 1140
Qy 1298 TGCTGCGTGGGAGCTTCAACAACCCCGCAAGAAACAACAGAGAGGCGCCCTTCTGT 1357
Db 1141 TGCTGCGTGGGAGCTTCAACAACCCCGCAAGAAACAACAGAGAGGCGCCCTTCTGT 1200
Qy 1358 GGAATGAGCTGAGAGCTGCAACCCCGCAAGTGAACCTGTGAGCTGAGAGCTGCGGAGA 1417
Db 1201 GGAATGAGCTGAGAGCTGCAACCCCGCAAGTGAACCTGTGAGAGCTGAGAGCTGCGGAGA 1260
Qy 1418 AGGAGAGCTGAGAGCTGTAACCAATCCAGAGAGCTGTGGCAAGCTGTAAGCTTGGCGAGCC 1477
Db 1261 AGGAGAGCTGAGAGCTGTAACCAATCCAGAGAGCTGTGGCAAGCTGTAAGCTTGGCGAGCC 1320
Qy 1478 AGATCTACCCCGGCAATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCC 1537
Db 1321 AGATCTACCCCGGCAATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCC 1380
Qy 1538 TGAACGACATGCTGTCCTTGAACGAGAGGCGGAGCTGAGAGCTGAGAGAGCCGCGAGA 1597
Db 1381 TGAACGAGGTATTCCTCTGACGAGAGGCGGAGCTGAGAGCTGAGAGAGCCGCGAGA 1440
Qy 1598 TCTGCGGAGAGCTGCGTGCACCGCGGTGTAAGTGAACCCCGCAAGAGACTGTGGCGGAGA 1657
Db 1441 TCTGCGGAGAGCTGCGTGCACCGCGGTGTAAGTGAACCCCGCAAGAGACTGTGGCGGAGA 1500
Qy 1658 TCCAGAGCAGGCGCGAGCAGAGCTGAGAGCTGAGAGCTGAGAGAGCTTCAAGAAACC 1717
Db 1501 TCCAGAGCAGGCGCGAGCAGAGCTGAGAGCTGAGAGCTGAGAGAGCTTCAAGAAACC 1560
Qy 1718 TGAAGACCGGCAAGTACCGCAAGATGAGCAACCGCGCAACCAAGAGCTGAGAGAGCTGAG 1777
Db 1561 TGAAGACCGGCAAGTACCGCGAGATGAGCAACCGCGCGCAACCAAGAGCTGAGAGAGCTGAG 1620
Qy 1778 CCGAGGCGGTGAGAGAGTCCGCAATGAGAGAGTGTGATCTGGGCGAAGACCCCGCAAGT 1837
Db 1621 CCGAGGCGGTGAGAGAGTCCGCAATGAGAGAGTGTGATCTGGGCGAAGATCCCGCAAGT 1680
Qy 1838 TCCGCGCTGCGCATCCAGAGAGAGACTGAGAGAGACTGAGAGAGCTGAGAGAGCTGAGAGAGCA 1897
Db 1681 TCCGCGCTGCGCATCCAGAGAGAGACTGAGAGAGACTGAGAGAGCTGAGAGAGCTGAGAGAGCA 1740
Qy 1898 CCGTGAATCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTAAGCTGTGATCAAGC 1957
Db 1741 CCGTGAATCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTAAGCTGTGATCAAGC 1800
Qy 1958 TGAAGAGAGGCTCATCTGAGCGCGAGAGACTTCTAGTGAACGCGCGCGCAACCGCG 2017
Db 1801 TGAAGAGAGGCTCATCTGAGCGCGAGAGACTTCTAGTGAACGCGCGCGCAACCGCG 1860
Qy 2018 AGAAGAGATGGGCAAGGCGGCTTACGTGAACCGAGCGAGGCGGCGAGAGATCTGAGGCG 2077
Db 1861 AGAAGAGATGGGCAAGGCGGCTTACGTGAACCGAGCGAGGCGGCGAGAGATCTGAGGCG 1920
Qy 2078 TGAAGAGAGCAACCAAGAGAGAGCGAGCTGAGAGAGCTTCAAGTGTGCGCTTGAAGAGCA 2137
Db 1921 TGAAGAGAGCAACCAAGAGAGAGCGAGCTGAGAGAGCTTCAAGTGTGCGCGCTTGAAGAGCA 1980
Qy 2138 GCGGCGAGAGGTGAACATCTGAGAGCGCAAGCTGAGAGCTTGGGCGATCTCAAGGCGCC 2197
Db 1981 GCGGCGAGAGGTGAACATCTGAGAGCGCAAGCTGAGAGCTTGGGCGATCTCAAGGCGCC 2040

Oy		2198	AGCCGACAAAGAGGCAAGCGACTGGTGAACCAGATCATGAGCAGCTGTACAAGAAG	2257
Db		2041	AGCCCGAACAAAGGAGGAGCGAGCTGGTGAGCCAGATCATGAGCAGCTGTACAAGAAG	2100
Oy		2258	AGAAGGTGACTTAGAGCTGGGTGCCTCCGCCAACAGGGGATCGGGGGCAAGAGAGATTCG	2317
Db		2101	AGAAAGGTGACTTAGCTGCTGGGTGCCTCCGCCAACAGGGGATCGGGGGCAAGAGAGATTCG	2160
Oy		2318	ACAAGCTGTGTAGCAAGGGGCAATCCGCGAAGTGTCTTTCTTGAAGCGGATCGATGGCGCA	2377
Db		2161	ACAAGCTGTGTAGGCGCGGCGATCCGCGAAGTGTCTTTCTTAAGCGGATCGATGGCGGCA	2220
Oy		2378	TCTGTATCTAACCAAGTACATGAGACGACTGTACGTGGGCAAGCGGGCCCTTAGATTCATT	2437
Db		2221	TCTGTATCTTACAGTACATGTAGACGACTGTACGTGGGCAAGCGGGCCCTTAGATTCATT	2280
Oy		2438	AAAAGCTTCCGGGGGCTAGCACCGGTGATAATC	2469
Db		2281	AAAAGCTTCCGGGGGCTAGCACCGGTGATAATC	2312
RESULT 5				
COS70576				
LOCUS		COS70576	2312 bp	DNA linear PAT 13-SEP-2004
DEFINITION		Sequence 84 from Patent EP143851.		
ACCESSION		COS70576		
VERSION		COS70576.1	GI:5200092	
KEYWORDS		.		
SOURCE		synthetic construct		
ORGANISM		synthetic construct		
REFERENCE		other sequences; artificial sequences.		
AUTHORS		1		
TITLE		Barnett,S., Zurnmagede,J., Sriyastava,I., Lian,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C. Improved expression of HIV polypeptides and production of virus-like particles		
JOURNAL		Patent: EP 1433851-A 84 30-JUN-2004; CHIRON CORPORATION (US)		
FEATURES		Location/Qualifiers		
source		1..2312		
		/organism="synthetic construct"		
		/mol_type="unassigned DNA"		
		/db_xref="taxon:32630"		
		/note="Description of Artificial Sequence: FS(-).promod.Rtopt(+)"		
ORIGIN				
		Query Match	83.1%; Score 2052;	DB 6; Length 2312;
		Best Local Similarity	93.6%; Pred.No.4,9e-168;	
		Matches 2165; Conservative	0; Mismatches 135; Indels	12; Gaps 2;
Oy		170	GCGGCAAGAGGGCCACCAAGATGAAGACTGACCGAGCGCGCAACTCTTCGCGG	229
Db		1	GCGGCGCGGAAGGACCACAATGAAAGATTGACTGAAGACAGGCTAATTTCTTCGCGG	60
Oy		2230	AGGACTTGAGCTTCCCTCCAGGGGCAAGCGCCGAGATTCCCTCCAGGAGCAAGACCGCGCA	289
Db		61	AGGACTTGAGCTTCCCTTGACAGGCGCAAGCGCCGAGATTGACAGCGCAGACCGCGCGCA	120
Oy		2290	ACAAGCCCCACGACCGCGGAGCTGCAAGTGCAGCGGCG-----ACAACCCCGCGACGAGG	343
Db		121	ACAAGCCCCACCGCGCGGAGCTGCAAGTGTGGGGCGCGAGAACAAAGCTGAGCGAGG	180
Oy		344	CCGGGCGCGGAGCGCCAGGCGAACCTCG-----AATCTCCCGCAATCACTGTGGCAGC	397
Db		181	CCGGGCGCGGAGCGCCAGGCGCACCGTGAAGCTTCAACTCTCCCGAATCACTGTGTGGCAGC	240
Oy		398	GCCCCCTGTGTGACATCAAGTGTGGGCGGCGCAGATCAAGAGGCGCTGTGTGACACCGGCG	457
Db		241	GCCCCCTGTGTGACATCAAGATTCGGGCGGCGCACTCAAGAGGCGGCTGTGTGACACCGGCG	300
Oy		458	CCGACGACACCGTGTGTGAGAGATGAGCTTCCCGCGCAAGTGAAGCCCAAATGATTCG	517

D	301	CCGACGACACCCGTCGTGGAAGATGAACCTGCCCCGCGCAAGTGGAAACCCAAATGATGTCG	360
Q	518	GCGGCATCCGGGCGCTTCATCAAGGTGCGCCAGTACGACCAAGATCTGATGAGATCTTGGC	577
D	361	GCGGGATCCGGGGGCTTTCATCAAGGTGCGGCAAGTACGACCAAGATCCCGTGAAGATCTTGGC	420
Q	578	GCAAGAAAGGCATTTGGGCACCGTGTGATTCGGCCCCACCCCGTGAACATCATTCGGCCGCA	637
D	421	GCCCAAGAGCCATGTGGCACCGTGTGTGGGCCCCCAACCCCGTGAACATCATTCGGCCGCA	480
Q	638	ACATGTGACCCAGCTGGGCTGCAACCCCTGAACCTTCCCCTACAGGCCCATCGAGACCGTGC	697
D	481	ACCTGTGACCCAGATCGGCTGCAACCCCTGAACCTTCCCCTACAGGCCCATCGAGACCGTGC	540
Q	698	CCGTGAAGCTGAAGCCCGGCAATGGAACGAGCCCAAGGTGAAGCACTGGCCCCCTGACCGAGG	757
D	541	CCGTGAAGCTGAAGCCCGGGAATGGAACGAGCCCAAGGTTCAGCACTGGCCCCCTGACCGAGG	600
Q	758	AGAAATTCAAAGGCCCTTGATCGGCCATCTTGCAGAGAGATGTGAAGAGAGAGGCAAGTTCACA	817
D	601	AGAAATTCAAAGGCCCTTGATCGGCCATCTTGCAGAGAGATGTGAAGAGAGAGGCAAGTTCACA	660
Q	818	AGATCGGCCCCGAGAACCCCTTACAAACAACCCCGGTTCGCCATCAAGAGAGAGAGACAGCA	877
D	661	AGATCGGCCCCGAGAACCCCTTACAAACAACCCCGGTTCGCCATCAAGAGAGAGAGACAGCA	720
Q	878	CCAAGTGCAGCAAGCTGATGATCTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTTGGG	937
D	721	CCAAGTGCAGCAAGCTGATGATCTTCGCGAGCTGAACAAAGCGCACCCAGGACTTCTTGGG	780
Q	938	AGGTGCAGCTGGGCATCCCCCAACCCCGCGCTGAAAGAGAGAGAGAGCTGACCGTGC	997
D	781	AGGTGCAGCTGGGCATCCCCCAACCCCGCGCTGAAAGAGAGAGAGAGCTGACCGTGC	840
Q	998	TGGAACGTGGGGGAAGGCTTACTTCAAGCGTGGCCCCCTGGAACAGAGACTTCGCGAAGTACACCG	1057
D	841	TGGAACGTGGGGGAAGGCTTACTTCAAGCGTGGCCCCCTGGAACAGAGACTTCGCGAAGTACACCG	900
Q	1058	CCTTCACCATCCCGAGCATCAACAACAGAACCCCGGATCCGCTACCAAGTACAGCTGC	1117
D	901	CCTTCACCATCCCGAGCATCAACAACAGAACCCCGGATCCGCTACCAAGTACAGCTGC	960
Q	1118	TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGATGAACCAAGTCTTGG	1177
D	961	TGCCCCAGGGCTGGAAGGGCAGCCCCAGCATCTTCCAGAGCAGATGAACCAAGTCTTGG	1020
Q	1178	AGCCCTTCCGCGCCCGCAACCCCGAGATCGTGAATCTTCAACTATATGGAACGACCTGTACG	1237
D	1021	AGCCCTTCCGCGAGAGAACCCCGACATCGTGAATCTTCAACTATATGGAACGACCTGTACG	1080
Q	1238	TGGGCAAGGCACTGGAGATCGGACAGCACGCGCCAGATGAGAGAGCTGGGCAAGCACCC	1297
D	1081	TGGGCAAGGCACTGGAGATCGGACAGCACGCGCCAGATGAGAGAGCTGGGCAAGCACCC	1144
Q	1298	TGCTGCGCTGAGGCTTCAACAACCCCGACAAAGAACACAGAGAGAGCCCCCTTCTGT	1357
D	1141	TGCTGCGCTGAGGCTTCAACAACCCCGACAAAGAACACAGAGAGAGCCCCCTTCTGT	1200
Q	1358	GGAATGGGCTTAAGAGTTGCAACCCCGACAAAGTGAACCGTGCACCCCATGAGCTGCCGAGA	1417
D	1201	GGAATGGGCTTAAGAGTTGCAACCCCGACAAAGTGAACCGTGCACCCCATGAGCTGCCGAGA	1260
Q	1418	AGAGAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGAGGCACGCC	1477
D	1261	AGAGAGCTGGAACCGTGAACGACATCCAGAAGCTGTGGGCAAGCTGAACCTGAGGCACGCC	1320
Q	1478	AGATTTACCCCGGCAATCAAGGTGCGCCAGCTGTGCAGACTGTCGCGCGGCCCAAGGCC	1537
D	1321	AGATTTACCCCGGCAATCAAGGTGGAAGCAGCTGTGCAGACTGTCGCGCGGCCCAAGGCC	1380
Q	1538	TGACCGGACATGTCGCTTGAACCGAGAGAGGCGGAGCTGGAACCTTGGCCGAGAACCGCGAGA	1597
D	1381	TGACCGGAGTATATCCCTTGAACCGAGAGAGGCGGAGCTGGAACCTTGGCCGAGAACCGCGAGA	1440

Db	841	TGAGACGTGGGGGAGAGCGCTACTTTCAGCGCGGCCCTTGAGCAGAGACTTCCGACAGTACACCG	900
OY	1058	CCCTTACCATCCTCCAGCATCAACAAGAGACCCCGGATCCGGTACCAGTACAAACGTGC	1117
Db	901	CCTTCAACATCTCCAGCATCAACAAGAGACCCCGGATCCGGTACCAGTACAAACGTGC	960
OY	1118	TGCCCCAAGGGCTGGAAAGGGCAGGCCCAAGCATCTTTCAGAGAGCATAGACCAAGTCTCTGG	1177
Db	961	TGCCCCAAGGGCTGGAAAGGGCAGGCCCGCATCTTTCAGAGAGCATAGACCAAGTCTCTGG	1020
OY	1178	AGCCCTTCCGCGCCCGCAACCCCGAGATCGATCTACCAATGATGAGAGCACTGTACG	1237
Db	1021	AGCCCTTCCGCAAGAGCAAGACCCCGAGATCTGTATCTACAGTACATAGAGCACTGTACG	1080
OY	1238	TGGGCAAGCGAAGCTGAGAGATTCGGCCAGCAACCGCGCCAGATTCAGAGAGCTGGCAAGACCC	1297
Db	1081	TGGGCAAGCGAAGCTGAGAGATTCGGCCAGCAACCGCACCAAGATTCAGAGAGCTGGCCAGACCC	1140
OY	1298	TGCTGCGCTGGGGGCTTTCACACACCCCGCAGAAAGACAGAGAGAGCCCCCTTCTGT	1357
Db	1141	TGCTGCGCTGGGGGCTTTCACACACCCCGCAGAAAGACAGAGAGAGCCCCCTTCTGT	1200
OY	1358	GGATGAGGCTTAGAGACTGACACCCCGACAGATGAGACCGTGCAGCCCATGAGCTGCGCCAGAG	1417
Db	1201	GGATGAGGCTTAGAGACTGACACCCCGACAGATGAGACCGTGCAGCCCATATGCTGCCAGAG	1260
OY	1418	AGGAGAGCTGAGACCGTGTAAACGACATCCAGAGTGTGTGGGCAAGCTGAACTGGGCGACGC	1477
Db	1261	AGGAGACGTTGAGACCGTGTAAACGACATCCAGAGTGTGTGGGCAAGCTGAACTGGGCGACGC	1320
OY	1478	AGATCTACCCCGGCAATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCC	1537
Db	1321	AGATCTACCGCGGCAATCAAGGTGTAAAGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCC	1380
OY	1538	TGACCGGACATGCTGCCCCCTGTACCGAGAGGCGCGAGCTGGAGACTGGCCGAGAACCGCGAGA	1597
Db	1381	TGACCGAGGTATCTCCCTGTACCGAGAGGCGCGAGCTGGAGACTGGCCGAGAACCGCGAGA	1440
OY	1598	TCCTGCGCGAAGCCCGTGTGACGCGCGTGTACTACAGACCCCGAGAGAGACTGTGTGCGCAGAG	1657
Db	1441	TCCTGTAAAGAGACCCGTTGACAGAGGTGTACTACAGACCCCGAGAGAGACTGTGTGCGCAGAG	1500
OY	1658	TCGAGAGCAGAGGCCAGACGACAGTGTGACCTTACCAAGTCTTACAGAGGCCCTTCAAGAAC	1717
Db	1501	TCGAGAGCAGAGGCCAGAGGCGAGTGTGACCTTACCAAGTCTTACAGAGGCCCTTCAAGAAC	1560
OY	1718	TGAAAGACCGGCAAGTACCGCCCAAGATGTGCGCACCGGCCACACCAAGAGCTGAAAGAGAGCTGA	1777
Db	1561	TGAAAGACCGGCAAGTACCGCCCGCAAGTGTGCGCGGCCCCACCAAGAGCTGAAAGAGAGCTGA	1620
OY	1778	CCGAGAGCGGTGCAAGAGATTCGCGCATGTGAGAGCATGTATCTGGGGCAGAGCCCCCAAGT	1837
Db	1621	CCGAGAGCGGTGCAAGAGGTGAGACCCGAGAGCATGTATCTGGGGCAGAGTCCCCCAAGT	1680
OY	1838	TCGCGCTGCGCATCCAGAGAGAGACTTGTGGAGACCTGTGTGACCGACTTACCTGTGAGCGCA	1897
Db	1681	TCAGGCTGCCATCCAGAGAGAGACTTGTGGAGGCGCTGTGTGATGTGAGTACTGTGACGCGCA	1740
OY	1898	CCTGATCTCCCGAGTGTGGAGATTCTGTGAACACCCCGCCCTGTGTGAAGCTGTGTGACACGC	1957
Db	1741	CCTGATCTCCCGAGTGTGGAGATTCTGTGAACACCCCGCCCTGTGTGAAGCTGTGTGACACGC	1800
OY	1958	TGAGAGAGAGCCATTCATTCGCGCGCGAGACCTTCTTACGTGTGACGCGCGCCGCAACCGCG	2017
Db	1801	TGAGAGAGAGCCATTCGTGGGCGCGAGACCTTCTTACGTGTGACGCGCGCCGCAACCGCG	1860
OY	2018	AGACCAAGATCGGCAAGGCTGGCTTACGTGACCGACCGGCGCGGAGAGATCTGTAGCC	2077
Db	1861	AGACCAAGTGTGGCAAGGCGGCTTACGTGACCGACCGGCGCGGAGAGATGTGTGAGCA	1920
OY	2078	TGACCGGAGACCAACCAAGAGAGACCGAGCTGACAGCCATTCAGTGTGCGCCCTTGCAGAGCA	2137
Db	1921	TCGCGGAGACCAACCAAGAGAGACCGAGCTGACAGCCATTCCTTGGCCCTTGCAGAGCA	1980

OY		2138	GGGCGACAGAGTGAACAATGTGTAAACCCGACAGCAAGTAAGGCCCTGGGGCATATCAGAGCCC	2197
Db		1981	GCGGCCTGTGAGGTGAACATGTGTACCCGACAGCCAAGTAGCCCTGGGGCATATCAGAGCCC	2040
OY		2198	AGCCCGACAAGAGGAGAGAGGAGCGCTGGTGTAACCAAGATCATTCGAGCAGCTGATCAAGAAGG	2257
Db		2041	AGCCCGACAAGAGGAGAGAGGAGCGCTGGTGTAACCAAGATCATTCGAGCAGCTGATCAAGAAGG	2100
OY		2258	AGAGGTGTACTTGAAGCTGGGTGCCCCGCCAACAAGGCGATCGGGGCGAACAGACAGATCG	2317
Db		2101	AGAGGTGTACTTGAAGCTGGGTGCCCCGCCAACAAGGCGATCGGGGCGAACAGACAGATCG	2160
OY		2318	ACAAGCTGTGTAGAGAAAGGAGATCCGCAAGGTGCTGTTCTTGGAACGGATCGATTGGGGGCA	2377
Db		2161	ACAAGCTGTGTAGAGCGCGGATCCGCAAGGTGCTGTTCTTGGAACGGATCGATTGGGGGCA	2220
OY		2378	TCTGTATCTACAGTATCATGGAACGACTGTACGTGGGACAGCGCGGCTTAGATTCATT	2437
Db		2221	TCTGTATCTACAGTATCATGGAACGACTGTACGTGGGACAGCGCGGCTTAGATTCATT	2280
OY		2438	AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC	2469
Db		2281	AAAAGCTTCCCGGGGCTAGCACCGGTGAATTC	2312
<hr/>				
RESULT 7				
LOCUS	AX427930	9166 bp	DNA	linear PAT 20-JUN-2002
DEFINITION	Sequence 168 from Patent WO0232943.			
ACCSSION	AX427930			
VERSION	AX427930.1 GI:21538017			
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE	1			
AUTHORS	Huang, Y. and Nabel, G.J.			
TITLE	Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization			
JOURNAL	Patent: WO 0232943-A 168 25-APR-2002;			
FEATURES	GOVERNMENT OF THE UNITED STATES (US)			
source	Location/Qualifiers 1..9166 /organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="Plasmid pVR1012x/s containing HIV genes"			
<hr/>				
ORIGIN				
<hr/>				
Query Match	83.1%; Score 2051.2; DB 6; Length 9166;			
Best Local Similarity	92.6%; Pred. No. 3.6e-168;			
Matches 2192; Conservative	0; Mismatches 163; Indels 13; Gaps 3;			
OY	14	TGGCGCAGAGCCATGAGCGAGGCGCACAGCGCCACATCTGTATGCAAGCGACGCAACTTCA	73	
Db	2961	TGGCGCAGAGCCATGAGCGAGGCGCACAGCGCCACATCTGTATGCAAGCGACGCAACTGCA	3020	
OY	74	AGGGGCCCAAGCGCATCATCAATGCTTTCMACTCGCGGCAAGAGGGGCGCATCTGCCCCGCA	133	
Db	3021	AGGGGCCCAAGCGCATCATCAATGCTTTCMACTCGCGGCAAGAGGGGCGCATCTGCCCCGCA	3080	
OY	134	ACTGCGCGGCCCCCGCGCAAGAAAGGGCTGTGGAAGTGCGGCGCAAGAGGGCCACGAGATGA	193	
Db	3081	ACTGCGCGGCCCCCGCGCAAGAAAGGGCTGTGGAAGTGCGGCGCAAGAGGGCCACGAGATGA	3140	
OY	194	AGGACTGCACCGAGCGGCCCAACTTCTTTCGCGAGAGCTTG3CTTTCGCCAGG3CA	253	
Db	3141	AGGACTGCACCGAGCGAGCAAGGCTTAA-TTTTTTAAAGGAAGACTTG3CTTTCGCCAGAG3CA	3199	
OY	254	AGGCGCGGAGATTTCCCGACGAGAGAAACCGCGCCACAGGCCACCAAGCGCGAGGCTGC	313	
Db	3200	AGGCGCGGAGATTTTCTTCAAGAGCAAGACAGAGCCCAAGCCCAACGAAGAGAGGCTTC	3259	

Db 4755 CCGAGAGCATCTGTGATCTGGGCGCAAGACCCCAAGTTCAAGCTGCCATCCAGAAAGAGA 4814
Qy 1862 CCTGGAGAGCCTGTGTGACCGA CTA CTGGCAGAGGCACTGGATCCCGAGTGGAGTTG 1921
Db 4815 CTTGGAGAGCCTGTGTGACCGAAGTACTGGACAGGCACTGGATCCCGAGTGGAGTTG 4874
Qy 1922 TGAACACCCCTGTGTGAACTGTGTGACCACTGGAGAAAGAGCCATCATCGAG 1981
Db 4875 TGAACACCCCTGTGTGAACTGTGTGACCACTGGAGAAAGAGCCATCATCGAG 4934
Qy 1982 CCGAAGCTTTTACGTGTGACCGGCGCCGCAACCGGAGACCAAGATCGGCAAGCGGCT 2041
Db 4935 CCGAAGCTTTTACGTGTGACCGGCGCCGCAACCGGAGACCAAGATCGGCAAGCGGCT 4994
Qy 2042 AGGTGACGACCGGCGCCGCAAGAGATGTTGAGCTTGACCGAGACCAACCAAGAGA 2101
Db 4995 AGGTGACGACCGGCGCCGCAAGAGTGTGCTTGTACCAACCAACCAAGAGA 5054
Qy 2102 CCGAGCTGAGGCGCATCCAGCTGCGCTTGACAGACAGGCGAGCGAGTGAACATCTGTA 2161
Db 5055 CCGAGCTGAGGCGCATCCAGCTGCGCTTGACAGACAGGCGCGTGGAGTGAACATCTGTA 5114
Qy 2162 CCGAAGCTTTACGCTGCGCTGCGCTGATCATCGAGCCGCAAGAGAGAGCGAGC 2221
Db 5115 CCGAAGCTTTACGCTGCGCTGCGCTGATCATCGAGCCGCAAGAGAGAGCGAGC 5174
Qy 2222 TGGTGAACGATCATCGAGCTGATGATGAGAGAGAGTGTACTGATGCTGGTGC 2281
Db 5175 TGGTGAACGATCATCGAGCTGATGATGAGAGAGAGTGTACTGATGCTGGTGC 5234
Qy 2282 CCGCCCAAGAGGCGCATCGCGGCAAGAGAGATCGAAGAGTGTGAGCAAGGCGATCC 2341
Db 5235 CCGCCCAAGAGGCGCATCGCGGCAAGAGAGATCGAAGAGTGTGAGCGGCGGCGATCC 5294
Qy 2342 GCAAGGTGCTGTTCTTCTGACCGGATGATGCGGCGATCTGATCTAACGTA 2393
Db 5295 GCAAGGTGCTGTTCTTCTGACCGGATGATGCGGCGCGCAAGAGAGAGCAAGAA 5346

RESULT 9
AX427931 9169 bp DNA linear PAT 20-JUN-2002
LOCUS AX427931
DEFINITION Sequence 169 from Patent WO232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS
TITLE
1 Huang Y. and Nabel G.J.
JOURNAL
genetic immunization
Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
source
1. 9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN
Query Match 82.2%; Score 2028.8; DB 6; Length 9169;
Beet Local Similarity 91.7%; Pred. No. 3.1e-166;
Matches 2194; Conservative 0; Mismatches 182; Indels 16; Gaps 4;
Qy 14 TGGCGAGGCGCATGAGCGAGCGCACCGAGCGCAACATCTGATGAGCGGAGCAACTTCA 73
Db 2967 TGGCGAGGCGCATGAGCGAGCGCACCGAGCGCAACATCTGATGAGCGGAGCAACTTCA 3026
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGCGCATCGCCGCA 133
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 3027 AGGGCCAGAGGCGCAT---CAAGTCTTCAACTGCGCGCAAGAGGCGCACTGCGCCGCA 3083
Qy 134 ACTGCGGCGCCCGCGCAAGAGGCGCTGTGAAATGTGGCGCAAGAGGCGCCACAGATGA 193
Db 3084 ACTGCGGCGCCCGCGCAAGAGGCGCTGTGAAATGTGGCGCAAGAGGCGCCACAGATGA 3143
Qy 194 AGAAGTGAACGAGGCGGAGGCGCAACTTCTTCGCGAGGACCTTGACCTTCCCGAGGCA 253
Db 3144 AGAAGTGAACGAGGCGGAGGCGCAACTTCTTCGCGAGGACCTTGACCTTCCCGAGGCA 3202
Qy 254 AGGCGCGGAGTTCCTCCAGCGAGCAACCGGCGCAACGCCCAACGCGCGAGCTGC 313
Db 3203 AGGCGCGGAGTTCCTTCAGAGCGAGCAAGGCGCAACGCCCAACGAGAGAGCTTC 3262
Qy 314 AGGTGCGGCGG-----CGAACAACCCCGGAGGCGCGGCGCGAGGCGCGAGCA--- 364
Db 3263 AGGTGCGGAGAGAGAGCAACACTCTCTCAAGAGCGAGGCGCGATGAGCAAGAGACTG 3322
Qy 365 ---CCCTGAACCTTCCCGAGATCACCTGTGGCAGCGCCCTGTGTGAGCATCAAGTGG 421
Db 3323 TATCTTTAGCTTCCCTGAGATCACTTTGGCAGCGACCTTGTTCATTAAGATAG 3382
Qy 422 GCGGCGAGATCAAGAGGCGCTGTGACACCGGCGCGAGCGACACCTGTGTGAGAGA 481
Db 3383 GCGGCGAGATCAAGAGGCGCTTCTAGACACCGGCGCGAGCGACACCTGTGTGAGAGA 3442
Qy 482 TGAAGCTTCCCGCGCAAGTGGAGAGCCCAAGATGAGTGGGCGCATCGGCGCTTCAACAAG 541
Db 3443 TGAAGCTTCCCGCGCGCTGGAGAGCCCAAGATGAGTGGGCGCATCGGCGCTTCAACAAG 3502
Qy 542 TGGCGCAATGACCAAGATCTGATCGAGATCTGCGGCAAGAGGCGCATCGGCGAGCTGC 601
Db 3503 TGGCGCAATGACCAAGATCTGATCGAGATCTGCGGCGCAAGAGGCGCATCGGCGAGCTGC 3562
Qy 602 TGATGCGCGCCACCCCGTGAACATCATCGCGCGCAACCTGTGACCAAGCTGGGCTGCA 661
Db 3563 TGATGCGCGCGCCACCCCGTGAACATCATCGCGCGCAACCTGTGACCAAGCTGGGCTGCA 3622
Qy 662 CCTGGAACCTTCCCGCATGAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGG 721
Db 3623 CCTGGAACCTTCCCGCATGAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGATGG 3682
Qy 722 AGGCGCCCAAGGTGAGAGTGGCCCTGTGACCGAGAGAGATCAAGGCGCTGACCGGCA 781
Db 3683 AGGCGCCCAAGGTGAGAGTGGCCCTGTGACCGAGAGAGATCAAGGCGCTGTGAGAGA 3742
Qy 782 TCTGCGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAACCTCTTACA 841
Db 3743 TCTGCGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAACCTCTTACA 3802
Qy 842 ACAACCCCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 901
Db 3803 ACAACCCCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 3862
Qy 902 TCCGAGAGCTGAAACAAGCGCACCGAGACTTGTGAGAGGTGAGCTGTGGCATTCGCCACC 961
Db 3863 TCCGAGAGCTGAAACAAGCGCACCGAGACTTGTGAGAGGTGAGCTGTGGCATTCGCCACC 3922
Qy 962 CCGCGGCGCTGAGAGAGAGAGAGAGCGGTGACCGGTGAGCGTGGCGGACGCTTACTTGA 1021
Db 3923 CCGCGGCGCTGAGAGAGAGAGAGAGCGGTGACCGGTGAGCGTGGCGGACGCTTACTTGA 3982
Qy 1022 GGTGCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 3983 GGTGCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4042
Qy 1082 AGAGAGACCCCGGAGCATCGCTTACAGTCAACAGTCTGCGCCGAGGAGTGAAGGCGAGCC 1141
Db 4043 AGAGAGACCCCGGAGCATCGCTTACAGTCAACAGTCTGCGCCGAGGAGTGAAGGCGAGCC 4102
Qy 1142 CCAAGCATTTTCAGAGAGAGAGAGAGAGATCTTGAAGAGCTTTCGCGCGCGCAACCCCG 1201
Db 4103 CCGCATCTTTCAGAGAGAGAGAGAGAGATCTTGAAGAGCTTTCGCGCGCGCAACCCCG 4162

Oy	1202	AGATCGGATCTTACACATACATGGAACAACGTGACCTGGGACAGACCTGGAGATGGCC	1261
Db	4163	ACATCGGATCTTACAGATACATGGAACAACGTGACCTGGGACAGACCTGGAGATGGCC	4222
Oy	1262	AGCACCGCGCCAGATCGAGGAGCTGGCGCAAGCACCTGTGCGTGGGCTTCAACAACC	1321
Db	4223	AGCACCGCACCAATGAGAGGAGCTGGCGCAAGCACCTGTGCGTGGGCTTCAACAACC	4282
Oy	1322	CCGACAGAGAGCACCAAGAGAGGCCCTTCTGTGGATGGGCTTACGAGCTGACCCCG	1381
Db	4283	CCGACAGAGAGCACCAAGAGAGGCCCTTCTGTGGATGGGCTTACGAGCTGACCCCG	4342
Oy	1382	ACAAGTGGACCGGTGCAGCCCATCCAGCTGCCCGAGAGAGAGCTGCACCGTGAACGACA	1441
Db	4343	ACAAGTGGACCGGTGCAGCCCATCCAGCTGCCCGAGAGAGAGCTGCACCGTGAACGACA	4402
Oy	1442	TCGAGAGCTGTGGGCAAGCTGAACCTGGACCGACCAAGATCTACCCGCACTCAAGGTGC	1501
Db	4403	TCGAGAGCTGTGGGCAAGCTGAACCTGGACCGACCAAGATCTACCGGCACTCAAGGTGC	4462
Oy	1502	GCCAGCTGTGAACCTGTGCGCGCGCCGAAAGGCCCTGACCGACATGATGCCCTGACCG	1561
Db	4463	GCCAGCTGTGAACCTGTGCGCGCGCACCAAGGCCCTGACCGAGTGTGCTCCCTGACCG	4522
Oy	1562	AGGAGGCCGAGCTGAGAGCTGGCCGAGAACCCGAGATCTTGCCGAGACCCGTGCACGCG	1621
Db	4523	AGGAGGCCGAGCTGAGAGCTGGCCGAGAACCCGAGATCTTGAAGAGACCCGTGCACGCG	4582
Oy	1622	TGTACTACGACCCCGACGAAAGGACTGTGGCCGAGATTCAGAGACGAGGCCACGACGAT	1681
Db	4583	TGTACTACGACCCCGACGAAAGGACTGTGGCCGAGATTCAGAGAACGAGGCCGACGAT	4642
Oy	1682	GGACCTTACGATCTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGA	1741
Db	4643	GGACCTTACGATCTTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCA	4702
Oy	1742	TGCGCACCGGCCCAACAACAAGCGTGAAGACCTGACCCGAGGCCGTGCAGAGATGCCCA	1801
Db	4703	TGAGAGGCGGCCCAACAACAAGCGTGAAGACGAGCTGACCCGAGGCCGTGCAGAGATGCCCA	4762
Oy	1802	TGGAGAGCATGTGTATCTGGGGGCAAGACCCCAAGTTCCGCTGACCATCCAGAGAGAGA	1861
Db	4763	CGAGAGCATGTGTATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAGAGAGA	4822
Oy	1862	CCTGGGAGACCTGGTGAACCGACTTACTGGCAGGCCACCTGTGATCCCGAGTGGAGTTCG	1921
Db	4823	CCTGGGAGAGCTGTGTGAACCGAGTACTGGCAGGCCACCTGTGATCCCGAGTGGAGTTCG	4882
Oy	1922	TGAACACCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGGCCCATCTACCTGGCG	1981
Db	4883	TGAACACCCCCCTGTGTGAAGCTGTGTGTACAGCTGGAGAGAGGCCCATCTACCTGGCG	4942
Oy	1982	CCGAGACCTTTCATCGTGTGAACGGCGCGCCCAACCGCGAGAACCAAGATCGGCAAGGCCCG	2041
Db	4943	CCGAGACCTTTCATCGTGTGAACGGCGCGCCCAACCGCGAGAACCAAGCTGGGACCGGCT	5002
Oy	2042	ACGTGACCGAACCGGGGCGGCGAAGAAAGTCTGTAGCGCTTGAACGAGACCAACAACAAGA	2101
Db	5003	ACGTGACCGAACCGGGGCGGCGAAGAGTGTGTGCTGTGACCGACCAACAACAAGA	5062
Oy	2102	CCGAGCTGCAAGCCCATTCAGAGCTGGCCCTTGCAGAGACAGCGGCGAGAGGTGAACATCGTGA	2161
Db	5063	CCGAGCTGCAAGCCCATTCAGAGCTGGCCCTTGCAGAGACAGCGGCTGTAGGTGAACATCGTGA	5122
Oy	2162	CCGACAGCAAGTACCGCCCTGGGACATCTACAGGCCCAACCCGACCAAGAGCGAGCGAGC	2221
Db	5123	CCGACAGCAAGTACCGCCCTGGGACATCTACAGGCCCAACCCGACCAAGAGCGAGCGAGC	5182
Oy	2222	TGTGTAAACAATCATTCGAGCAAGCTGTATCAGAGAGAGAGAGGTGTACTGTGACCTGGGTGC	2281
Db	5183	TGTGTAAACAATCATTCGAGCAAGCTGTATCAGAGAGAGAGAGGTGTACTGTGACCTGGGTGC	5242

QY	2282	CCGGCCCAACAGGGGATCGGGGGGCAAGAGATCGAACAGCTGTGAGCAAGGGCAATCC	2341
DB	5243	CCGCCCAACAGGGGATCGGGGCAAGAGATCGAACAGCTGTGAGCAAGGGCAATCC	5302
QY	2342	GCAAGTGTCTTCTCTGAGCGGCATCGATGGCGGCATCTGTATCTACCAAGTA	2393
DB	5303	GCAAGTGTCTTCTCTGAGCGGCATCGACAAAGGCCAGAGAGCAAGAA	5354
RESULT 10			
LOCUS	AX427925	9194 bp	DNA
DEFINITION	Sequence 163 from Patent WO0232943.	linear	PAT 20-JUN-2002
ACCESSION	AX427925		
VERSION	AX427925.1	GI:21538012	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	1		
TITLE	Huang,Y. and Nabel,G.U.		
JOURNAL	Modifications of hiv env, gag, and pol enhance immunogenicity for genetic immunization		
FEATURES	Patent: WO 0232943-A 163 25-APR-2002; GOVERNMENT OF THE UNITED STATES (US)		
source	Location/Qualifiers		
	1..9194		
ORIGIN	/organism="synthetic construct" /mol_type="unassigned DNA" /db_xref="taxon:32630" /note="plasmid pVR1012x/s containing HIV genes"		
Query Match	82.1%; Score 2027; DB 6; Length 9194;		
Best Local Similarity	91.6%; Pred. No. 4.4e-166;		
Matches 2194; Conservative	0; Mismatches 185; Indels 16; Gaps 4;		
QY	14	TGGCCGAGGCGCATGAGCCAGG---CCACAGCGGCAACATCTGTATGAGCGACGAAC	70
DB	2967	TGGCCGAGGCGCATGAGCCAGGCGGACCAAGCGGCAACATCATATATGAGCGCGGCAACT	3026
QY	71	TCAAGGGCCCCAAGGCGCATCATCAAGTGTCTTCAACTGGCGCAAGAGGGCCACATCGCC	130
DB	3027	TCCGCAACCAAGCGCAAGATCGTGAAGTGTCTTCAACTGGCGGCAAGAGGGCCACACCGCC	3086
QY	131	GCACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGTGCGCAAGAGGGCCACCAAGA	190
DB	3087	GCAATGTGCGCGCGCCCCCGCAAGAGGGCTGTGAAAGTGTGCGCAAGAGGGCCACCAAGA	3146
QY	191	TGAAGCATGTGACCGAGCGCGGCGCAACTCTTCGCGGAGACCTGTGCTTCCCGCAGG	250
DB	3147	TGAAGCATGTGACCGAGCGGCGGAGGCTTA-TTTTTTAAGGAAGATGTGGCTTTCCCAAG	3205
QY	251	GCAAGGCCCGGAGATTCCCGACGACGAGAAACCGCGCCACAGGCCCCACGACCGGAGC	310
DB	3206	GGAAGCGCAAGGGAATTTTCTTCAAGGCGAGACCAAGGCCACAGCCCCACGAAAGAGGC	3265
QY	311	TGCAGGTGCGCGG-----CGACAAACCCCGCGAGAGGCCGGCGCGGCGGCGGCGGCA	364
DB	3266	TTCAAGTGTGGGGAGAGAACAACTCCCTCTCGAAGACAGAGGCCGATAGCAAGGAA	3325
QY	365	-----CCCTGAACCTTCCCGCAAGATCAACCTGTGGCAGCGCGCCCTGTGTGAGCATCAAG	418
DB	3326	CTGTATCTCTTAACTTCCCTCAATCACTCTTTGGCAGCGACCCCTGTCAATTAAGA	3388
QY	419	TGGCGGCGCAATCAAGAGGCGCTGTGACACCGCGGCGCGACGACACCGTGTCTGAGG	478
DB	3386	TAGGGGGCGAAGTGAAGAGGCGCTTCAAGACACCGCGCGCGACGACACCGTGTCTGAGG	3445
QY	479	AGATAGCGCTGCGCGGCAAGTGAAGAGCCCAAGATGATGGGGGAGATGGCGCGCTTCAACA	538
DB	3446	AGATAGCGCTGCGCGCGCTGAAGAGCCCAAGATATATGGCGGAGATGGCGCGCTTCAACA	3506

539 AGTGGGCGGAGTACGACGAGATCTGTAGTGGAGATCTGGCGGAAAGAGCCGATCGGACCG 598
5306 AGTGGGCGGAGTACGACGAGATCTGTAGTGGAGATCTGGCGGAAAGAGCCGATCGGACCG 3565
539 TGTGTATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 658
3566 TGTGTATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3625
659 GCACCTGTAACTTCCCATCAGCGCCCATCAGACCGTGGCCCGTGAAGCTGAAGCCCGGCA 718
3626 GCACCTGTAACTTCCCATCAGCGCCCATCAGACCGTGGCCCGTGAAGCTGAAGCCCGGCA 3685
719 TGGACGGCCCGCAAGGTGAAGAGTGGCCCTTGAACCGAGGAAGATCAAGGCCCTGACCG 778
3686 TGGACGGCCCGCAAGGTGAAGAGTGGCCCTTGAACCGAGGAAGATCAAGGCCCTGACCG 3745
779 CCATCTGCGAGAGATGAGAGAGAGAGAGAGATCAACAGATCGGCGCCCGGAGAACCTT 838
3746 AGATCTGCGAGAGATGAGAGAGAGAGAGATCAACAGATCGGCGCCCGGAGAACCTT 3805
839 ACAACACCCCGGTGTCCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 898
3806 ACAACACCCCGGTGTCCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3865
899 ACTTCGCGAGCTGAAACAAGCGCACCCAGACTTCTGGAGGTGCACTGGGCAATCCCGC 958
3866 ACTTCGCGAGCTGAAACAAGCGCACCCAGACTTCTGGAGGTGCACTGGGCAATCCCGC 3925
959 ACCCGCGCGCGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
3926 ACCCGCGCGCGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3985
1019 TCAGCGTGGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1078
3986 TCAGCGTGGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4045
1079 ACAACAGAGACCCCGGATCGGCTACAGTACAGTGTGCTGCCAGAGGCTGAAAGAGAG 1138
4046 ACAACAGAGACCCCGGATCGGCTACAGTACAGTGTGCTGCCAGAGGCTGAAAGAGAG 4105
1139 GCCCGAGATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
4106 GCCCGAGATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4165
1199 CCGAGATGTGTATCTACAGTACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1258
4166 CCGAGATGTGTATCTACAGTACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4225
1259 GCCAGAGACCGCGCAAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318
4226 GCCAGAGACCGCGCAAGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4285
1319 CCCCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378
4286 CCCCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4345
1379 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1438
4346 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4405
1439 ACATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1498
4406 ACATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4465
1499 TGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1558
4466 TGGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4525
1559 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1618
4526 CCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4585
1619 GCGTGTACTACGACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1678

4586 GCGTGTACTACGACCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4645
1679 AGTGAACCTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1738
4646 AGTGAACCTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4705
1739 AGATGCGACCGCGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1798
4706 GATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4765
1799 CCATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
4766 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4825
1859 AGACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
4826 AGACCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4885
1919 TGTGAACACCCCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1978
4886 TGTGAACACCCCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4945
1979 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
4946 GCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5005
2039 GCTAGGTACCGAGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2098
5006 GCTAGGTACCGAGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5065
2099 AGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158
5066 AGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5125
2159 TGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2218
5126 TGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5185
2219 AGCTGTGATACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2278
5186 AGCTGTGATACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5245
2279 TGGCGGCGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338
5246 TGGCGGCGCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5305
2339 TCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2393
5306 TCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5360

RESULT 11
AX427927
LOCUS AX427927 12411 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 165 from Patent WO0232943.
ACCESSION AX427927
VERSION AX427927.1 GI:21538014
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G. J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
JOURNAL Patent: WO 0232943-A 165 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1..12411
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"

ORIGIN /note="plaemid pvr1012x/s containing HIV genes"

Query Match 82.1%; Score 2027; DB 6; Length 12411;
Best Local Similarity 91.6%; Pred. No. 4e-166;
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;

QY 14 TGGCGGAGGCGCTGAGGCCAGG---CCACGAGGCCCAACCTCTGATGCAAGCCAGCAACT 70
DB 2967 TGGCGGAGGCGCTGAGGCCAGGCGCCACCACTGATGCAAGCCAGCAACT 3026
QY 71 TCAAGGGCGCCCAAGGCGCATCATCAAGTGTCTCACTGCGGCGCAAGAGGGGCGCATTCGCCC 130
DB 3027 TCCGCAACCAAGCGCAAGATGTGAAGTGTCTCACTGCGGCGCAAGAGGGGCGCATTCGCCC 3086
QY 131 GCNACTGCGCGCGCGCGCGCAAGAGGGGCTGTGAAGTGTGCGGCAAGAGGGGCGCAACAGA 190
DB 3087 GCAACTGCGCGCGCGCGCGCAAGAGGGGCTGTGAAGTGTGCGGCAAGAGGGGCGCAACAGA 3146
QY 191 TGAAGGACTGCAACCGAGCGCGCAAGCCACTTCTTCCGCGAGGACTGTGCGCTTCCCGCAG 250
DB 3147 TGAAGGACTGCAACCGAGCGCGCAAGCCACTTCTTCCGCGAGGACTGTGCGCTTCCCGCAG 3205
QY 251 GGAAGGCGCGCGAGTTCGCCAGGAGCAAGCGGCGCAACGCGCCCAAGCGCGGAGC 310
DB 3206 GGAAGGCGCGAGAAATTTCTTCAAGAGAGACCAAGCGCAACGCGCCCAAGAGAGAGC 3265
QY 311 TGCAGGTGCGCGG-----CGACAACCGCGGAGCGAGGCGCGCGCGCGCGCAGGCGCA 364
DB 3266 TTCAAGGTTTGGGGAGAGACCAACAATCTCTCTCAAGAGAGAGAGCGCGCAAGAGAA 3325
QY 365 -----CCTGAACTTCCCGCAAGTCAACCTGTGCGAGCGCGCGCTGTGAAGATCAAG 418
DB 3326 CTGTATCTTTAGCTTCCCTCAGATCACTTTGGGAGCGACCGCTCTGCACAATTAAGA 3385
QY 419 TGGGCGCGCGAGTCAAGGAGGCGCTGTGAACAACGCGCGCGCAACGCGTGTGAAG 478
DB 3386 TTAGGGGCGCGAGTCAAGGAGGCGCTTCTAGACAACGCGCGCGCAACGCGTGTGAAG 3445
QY 479 AGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATTCGCGCGCATCGCGCGCTTCACTA 538
DB 3446 AGATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCACTA 3505
QY 539 AGGTGCGCGAGTCAAGCAAGTCTGATCGAGTCTGCGCGCAAGAGCGCATCGAGCA 598
DB 3506 AGGTGCGCGAGTCAAGCAAGTCTGATCGAGTCTGCGCGCAAGAGCGCATCGAGCA 3565
QY 599 TGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 658
DB 3566 TGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3625
QY 659 GCAACCTGAACCTTCCCATCAGCGCCCATGAGACCGTGCCTGTGAAGCTGAAGCCGCGCA 718
DB 3626 GCAACCTGAACCTTCCCATCAGCGCCCATGAGACCGTGCCTGTGAAGCTGAAGCCGCGCA 3685
QY 719 TGAAGCGCGCGCAAGTGAAGCAAGTGCCTTGAACCGAGAGAAAGATCAAGGCCCTGAC 778
DB 3686 TGAAGCGCGCGCAAGTGAAGCAAGTGCCTTGAACCGAGAGAAAGATCAAGGCCCTGAC 3745
QY 779 CCATCTGCGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 838
DB 3746 AGATCTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3805
QY 839 ACAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 898
DB 3806 ACAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3865
QY 899 ACTTCCGCGAGCTGAACAGAGCGCAACCGAGACTTCTGAGAGGTGCAAGTGGGCAATCCCC 958
DB 3866 ACTTCCGCGAGCTGAACAGAGCGCAACCGAGACTTCTGAGAGGTGCAAGTGGGCAATCCCC 3925
QY 959 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018

DB 3926 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3985
QY 1019 TCAAGGTGCGCGCTGTGAACAGAGACTTCCGCAAGTCAACCGCGCTTCAACATCCCGACATCA 1078
DB 3986 TCAAGGTGCGCGCTGTGAACAGAGACTTCCGCAAGTCAACCGCGCTTCAACATCCCGACATCA 4045
QY 1079 ACAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1138
DB 4046 ACAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4105
QY 1139 GCCCGAGACTTCTCAAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGCGCGCGCAAC 1198
DB 4106 GCCCGAGACTTCTCAAGAGCAGCATGACCAAGATCTGAGGCGCTTCCGCGCGCGCAAC 4165
QY 1199 CCGAGATGTGATCTTCAAGATCATGAGACCTTGTAGTGGGCGAGCGCGCGCGCGCGAG 1258
DB 4166 CCGAGATGTGATCTTCAAGATCATGAGACCTTGTAGTGGGCGAGCGCGCGCGCGCGAG 4225
QY 1259 GCGAGACCGCGCGCAAGATCGAGGAGCTGCGGAGACACTGTGCGCTGAGGCGCTTCAACA 1318
DB 4226 GCGAGACCGCGCAAGATCGAGGAGCTGCGGAGACACTGTGCGCTGAGGCGCTTCAACA 4285
QY 1319 CCGCGAGAGAGAGACCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1378
DB 4286 CCGCGAGAGAGAGACCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4345
QY 1379 CCGAGAGTGAACCGTGAACCGCATGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAG 1438
DB 4346 CCGAGAGTGAACCGTGAACCGCATGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAG 4405
QY 1439 AATATCAAGAGTGTGCGCGCAAGCTGAGGCGCGAGCGCGAGCTTCAACCGCGCATCAAG 1498
DB 4406 AATATCAAGAGTGTGCGCGCAAGCTGAGGCGCGAGCGCGAGCTTCAACCGCGCATCAAG 4465
QY 1499 TGCAGCACTGTGCAAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1558
DB 4466 TGCAGCACTGTGCAAGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4525
QY 1559 CCGAGAGGCGCGAGCTGAGCTGCGCGAGAGACCGCGAGATCTTCCGCGCGAGCGCG 1618
DB 4526 CCGAGAGGCGCGAGCTGAGCTGCGCGAGAGACCGCGAGATCTTCCGCGCGAGCGCG 4585
QY 1619 GGTGTACTAGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1678
DB 4586 GGTGTACTAGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4645
QY 1679 AGTGAACCTAAGAGATCTTCAAGAGCGCGCTTCAAGAGAGAGAGAGAGAGAGAGAG 1738
DB 4646 AGTGAACCTAAGAGATCTTCAAGAGCGCGCTTCAAGAGAGAGAGAGAGAGAGAGAG 4705
QY 1739 AGATGCGCACGCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1798
DB 4706 AGATGCGCACGCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4765
QY 1799 CCATGAGAGAGATGTGATCTTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1858
DB 4766 CCATGAGAGAGATGTGATCTTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4825
QY 1859 AGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1918
DB 4826 AGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4885
QY 1919 TGTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1978
DB 4886 TGTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4945
QY 1979 GCGCGAGAGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2038
DB 4946 GCGCGAGAGCTTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5005
QY 2039 GCTAAGTGAACCGAGCGCGCGCGCGAGAGATCTGAGCTGAGCGAGAGAGAGAGAGAG 2098
DB 5006 GCTAAGTGAACCGAGCGCGCGCGCGAGAGATCTGAGCTGAGCGAGAGAGAGAGAGAG 5065

QY 2099 AGACGAGCTGACGAGCCATTCAGCTGACCTGTCAGAGGACGAGCCAGAGGTGAACATCG 2158
DB 5066 AGACGAGCTGACGAGCCATTCAGCTGACCTGTCAGAGGACGAGCCAGAGGTGAACATCG 5125
QY 2159 TGACCGACGACGAGTACGCGCTGAGGATCATTCAGGCGCCAGCCGACAGAGGAGAGG 2218
DB 5126 TGACCGACGACGAGTACGCGCTGAGGATCATTCAGGCGCCAGCCGACAGAGGAGAGG 5185
QY 2219 AGCTGATGACGAGTACGAGTATTCAGGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGG 2278
DB 5186 AGCTGATGACGAGTACGAGTATTCAGGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGG 5245
QY 2279 TGCCGCGCCCAAGAGGACATCGAGCGACAGAGGATCGACAGAGCTGTGAGCAAGAGCA 2338
DB 5246 TGCCGCGCCCAAGAGGACATCGAGCGAGAGAGAGAGTGTGACCGCTGTGAGAGCGCGCA 5305
QY 2339 TCCTGCAAGGTGCTGTTCTCTGAGACGAGCATGATGAGCGAGCATCGTGTACTACAGTA 2393
DB 5306 TCCTGCAAGGTGCTGTTCTCTGAGACGAGCATGAGAGCGCCAGAGAGAGAGAGAG 5360

RESULT 12
AX427926 9194 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 164 from Patent WO0232943.
DEFINITION AX427926
ACCESSION AX427926
VERSION AX427926.1 GI:21538013
KEYWORDS
SOURCE
ORGANISM
1 synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
TITLE
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 164 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
1..9194
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR102x/s containing HIV genes"

ORIGIN
Query Match 82.0%; Score 2025.4; DB 6; Length 9194;
Best Local Similarity 91.6%; Pred. No. 66-166;
Matches 2193; Conservative 0; Mismatches 186; Indels 16; Gaps 4;
QY 14 TGGCCGAGCGCCATGAGCCAGG---CCACGAGCGCCAACTCTGTATGACGCGCAACT 70
DB 2967 TGGCCGAGCGCCATGAGCGCGAGGACCAACAGCGCCCACTCATGATGACGCGCAACT 3026
QY 71 TCAAGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCGCATGCGCC 130
DB 3027 TCCGCAACCAAGCGCAAGATCGTGAAGTCTTCACTGCGGCAAGAGAGGCGCATGCGCC 3086
QY 131 GCAACTGCGCGCGCCCGCCGCAAGAGAGGCTGCTGAAAGTGCAGAGAGGCGCAACAGA 190
DB 3087 GCAACTGCGCGCGCCCGCCGCAAGAGAGGCTGCTGAAAGTGCAGAGAGGCGCAACAGA 3146
QY 191 TGAAGACCTGACCGAGCGCGCGCAACTTCTTCCGCAAGAGACTGTGCGCTTCCCGCAG 250
DB 3147 TGAAGACCTGACCGAGCGCGCGCAAGGCTTA-TTTTATGAGAAATCTGCGCTTCCCAAG 3205
QY 251 GCAAGCGCGCGAGTTCCCGAGGAGAGACGAGCGCAAGCGCCCAAGCGCGAGC 310
DB 3206 GCAAGCGCGAGAAATTTTCTTCAAGAGCAACAGAGCGCCCAAGAGAGAGC 3265
QY 311 TGCAGGTGCGCG-----CGAACCCCGCAGCGAGCGCGCGCGCGCGCGCAAGGCA 364
DB 3266 TGCAGGTGCGCGAGAGAGCAACCACTCCTCTCAAGAGAGAGCGAGCGAGTAAAGAGAA 3325

QY 365 -----CCTGAACCTTCCCCCAGATCACCTGTGGCAGCGCCCGTGTGAGCATCAAG 418
DB 3326 CTGTATCTCTTACCTTCCCTCAGATCACTCTTGTGAGAGAGCCCTCGTCAAAATAAAGA 3385
QY 419 TGGGCGCCAGATCAAGAGAGCGCTGTGACACCGGCGCCAGACACCGTCTGAGAG 478
DB 3386 TAGGGGCGCAGCTGAAGAGAGCGCTTCTTAAACACCGGCGCCGAGACACCGTGTGAGAG 3445
QY 479 AGATGAGCTTGGCCCGCAAGTGAAGGCCCAAGATGATGGCGGCGATTCGGCGCTTCACTA 538
DB 3446 AGATGAGCTTGGCCCGCGCTGGAAGGCCCAAGATGATGGCGGCGATTCGGCGCTTCACTA 3505
QY 539 AGTGGCGCAGTACGACAGAGATCTGTATCGAGATCTGCGGCAAGAGAGCGATCGGACCG 598
DB 3506 AGTGGCGCAGTACGACAGAGATCTGTATCGAGATCTGCGGCAAGAGAGCGATTCGGACCG 3565
QY 599 TGTGATCGGCGCCACCGCCGTGAACATCATCGCGCAACATGCTGACCCAGCTGAGCT 658
DB 3566 TGTGATCGGCGCCACCGCCGTGAACATCATCGCGCAACATGCTGACCCAGATTCGCTT 3625
QY 659 GCACTCTGAACCTTCCCATGAGCCCATGAGACCGTGGCCGTGAAGCTGAAGCCCGCA 718
DB 3626 GCACTCTGAACCTTCCCATGAGCCCATGAGACCGTGGCCGTGAAGCTGAAGCCCGCA 3685
QY 719 TGGAGCGCCCAAGGTGAAGAGAGTGGCCCTGACCGAGAGAGAGATCAAGGCGCTGACCG 778
DB 3686 TGGAGCGCCCAAGGTGAAGAGAGTGGCCCTGACCGAGAGAGAGATCAAGGCGCTGAGTGG 3745
QY 779 CCATCTGAGAGAGATGAGAGAGAGAGAGAGATCAAGATTCGCGCCCGAGAACCCCT 838
DB 3746 AGATCTGACCGAGATGAGAGAGAGAGAGAGATCAAGATTCGCGCCCGAGAACCCCT 3805
QY 839 ACAACACCCCGGTGTTCCCATCAAGAGAGAGAGAGAGATCAAGATTCGCGCCCGAGAACCCCT 898
DB 3806 ACAACACCCCGGTGTTCCCATCAAGAGAGAGAGAGAGATCAAGATTCGCGCCCGAGAACCCCT 3865
QY 899 ACTTCGAGAGCTGAACAGAGAGAGAGAGAGATTCGAGAGAGTGAAGCTGAGCATCCCC 958
DB 3866 ACTTCGAGAGCTGAACAGAGAGAGAGAGAGATTCGAGAGAGTGAAGCTGAGCATCCCC 3925
QY 959 ACCCGCGCGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1018
DB 3926 ACCCGCGCGCTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3985
QY 1019 TCAAGTGGCCCTGAGACAGAGACTTCGCAAGTACACGCTTCAACCATCCCAAGATCA 1078
DB 3986 TCAAGTGGCCCTGAGACAGAGACTTCGCAAGTACACGCTTCAACCATCCCAAGATCA 4045
QY 1079 ACAAGAGACCCCGCGCATTCGCTACAGTACAGTGAAGAGTGGCCGAGAGAGAGAGAG 1138
DB 4046 ACAAGAGACCCCGCGCATTCGCTACAGTACAGTGAAGAGTGGCCGAGAGAGAGAGAGAG 4105
QY 1139 GCCCGAGACTTTCAGAGCAGACAGACCAAGATCTGTGAGGCTTTCGCGCGCCGCAAC 1198
DB 4106 GCCCGAGACTTTCAGAGCAGACAGACCAAGATCTGTGAGGCTTTCGCGCGCCGCAAC 4165
QY 1199 CGAGATGTGATCTTACAGTACATGACAGACCTGTAGCTGGCGAGCGACTGTAGATCG 1258
DB 4166 CGAGATGTGATCTTACAGTACATGACAGACCTGTAGCTGGCGAGCGACTGTAGATCG 4225
QY 1259 GCGACACCGCGCGCAAGATTCAGAGAGCTGGCGAACACCTGTGAGCTGGCGCTTCAACA 1318
DB 4226 GCGACACCGCGCGCAAGATTCAGAGAGCTGGCGAACACCTGTGAGCTGGCGCTTCAACA 4285
QY 1319 CCCCCGACGAAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1378
DB 4286 CCCCCGACGAAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4345
QY 1379 CCGACAGTGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1438
DB 4346 CCGACAGTGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4405

1439 ACATCCAGAGGCTGTGGGCAAGCTGAACTGGGCGACGACGATCTAACCCCGGCATCAAGG 1498
1406 ACATCCAGAGGCTGTGGGCAAGCTGAACTGGGCGACGACGATCTAACCCCGGCATCAAGG 1465
1499 TGGCGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCTGACCGACATCTGCTCCCTGA 1558
4466 TGGCGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCTGACCGACGATGCTGCTCCCTGA 4525
1559 CCGAGAGGCGGAGCTGTGAGCTGTGCGCGGACCGGAGATCTGCGCGAGCGCTGTGACG 1618
4526 CCGAGAGGCGGAGCTGTGAGCTGTGCGCGGACCGGAGATCTGAGAGAGCGCGGTGACG 4585
1619 GGGTGTACTACGACCCCGACGACGACCTGTGGCGGACGATCTGAGAGAGCGCGGACGAC 1678
4586 GGGTGTACTACGACCCCGACGACGACCTGTGGCGGACGATCTGAGAGAGCGCGGACGAC 4645
1679 AGTGGACCTACGAGATCTACCGAGAGCGCTTCAAGACCTGAGAGCGCGCAAGTACGCA 1738
4646 AGTGGACCTACGAGATCTACCGAGAGCGCTTCAAGACCTGAGAGCGCGCAAGTACGCCC 4705
1739 AGATGGGACCGGCGGCAACCAACGAGTGAAGGAGCTGAACGAGCGCTGTGCAAGAGTGC 1798
4706 GGAATGAGGCGGCGGCAACCAACGAGTGAAGGAGCTGAACGAGCGCGGTGCAAGAGTGC 4765
1799 CCAATGAGAGCTGTGATCTGGGCGCAAGACCCCGAAGTTCGCGCTGCGCATTCAGAGG 1858
4766 CCAATGAGAGCTGTGATCTGGGCGCAAGACCCCGAAGTTCAGAGCTGCGCATTCAGAGG 4825
1859 AGACCTGGGAGAGCTGTGAGACCGACCTAGGAGGCGCACTGGATCCCGAGTGGAGAT 1918
4826 AGACCTGGGAGAGCTGTGAGACCGACCTAGGAGGCGCACTGGATCCCGAGTGGAGAT 4885
1919 TGTGTAACACCCCGCTGTGTGAAGCTGTGTGACGAGTGAAGAGAGCCCATCATCG 1978
4886 TGTGTAACACCCCGCTGTGTGAAGCTGTGTGACGAGTGAAGAGAGCCCATCATCG 4945
1979 GGGCGGAGACCTTCTAGTGAAGCGCGCGGCAACCGCGGACCAAGATCGGCAAGCGCG 2038
4946 GGGCGGAGACCTTCTAGTGAAGCGCGCGGCAACCGCGGACCAAGATCGGCGCAAGCGCG 5005
2039 GCTAGGTAACGACCGGCGGCGGCAAGATGTGTGAGCTGACCGAGACCAACCAACGAG 2098
5006 GCTAGGTAACGACCGGCGGCGGCAAGATGTGTGAGCTGACCGAGACCAACCAACGAG 5065
2099 AGACCGAGCTGACGAGCTGACGAGTGCCTGTGAGGACGAGCGGACGAGTGAATCG 2158
5066 AGACCGAGCTGACGAGCTGACGAGTGCCTGTGAGGACGAGCGGCTGTGAGTGAATCG 5125
2159 TGAACGACGAGCTGACGAGCTGAGCTGAGCTGACGAGCTGACGAGCTGACGAGCTGACG 2218
5126 TGAACGACGAGCTGACGAGCTGAGCTGAGCTGACGAGCTGACGAGCTGACGAGCTGACG 5185
2219 AGCTGTGAGACGAGCTGACGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 2278
5186 AGCTGTGAGACGAGCTGACGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 5245
2279 TGGCGCGGCAAGGCGCATGCGCGGCAACGAGGAGTGAACGAGCTGTGTGAGGAGGCGCA 2338
5246 TGGCGCGGCAAGGCGCATGCGCGGCAACGAGGAGTGAACGAGCTGTGTGAGGAGGCGCGCA 5305
2339 TCGGCAAGGCTGTGCTTCTGAGCGGCACTGAGCGGCGCATGTGATCTACAGTA 2393
5306 TCGGCAAGGCTGTGCTTCTGAGCGGCACTGAGCGGCGCATGTGATCTACAGTA 5360

RESULT 13
LOCUS BD263704 2306 bp DNA linear PAT 17-JUL-2003
DEFINITION Improved expression of HIV polypeptides and production of
virus-like particles.
ACCESSION BD263704.1 GI:33073472
VERSION BD263704.1
KEYWORDS JP 200253124-A/71.

SOURCE
ORGANISM synthetic construct
REFERENCE
AUTHORS 1 (bases 1 to 2306)
Barnett, S., Megede, J. Z., Srivastava, I., Lian, Y., Hartog, K., Liu, H., Greer, C., Selby, M. and Walker, C.
TITLE Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: JP 200253124-A 71 08-OCT-2002;
COMMENT
CHIRON CORP
OS Artificial Sequence
PN JP 200253124-A/71
PD 08-OCT-2002
PR 30-DEC-1999 JP 2000591193
PI 60/114495, 01-DEC-1999 US 60/168471
PI SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI KARI HARTOG,
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER
PC C12N15/09, A61K31/711, A61K38/00, A61K46/00, A61P31/18, A61P37/02, PC C12N5/10,
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
Description of Artificial Sequence: FS(-). protomod. RTopl. YM FH
FT source 1. .2306
Location/Qualifiers
FEATURES
source 1. .2306
Location/Qualifiers
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
170 GCGGCAAGAGGCGCAACGATGAAGGAGTGAACCGAGGCGGAGCGCACTTCTGCGCG 229
1 GCGGCGGCAAGAGGCGCAACGATGAAGGAGTGAACCGAGGCGGAGCGCGCACTTCTGCGCG 60
230 AGGACCTGCGCTTCCCGGCAAGGCGGCGGAGTTCGCCAGCGAGCAACCGCGCA 289
61 AGGACCTGCGCTTCCCGGCAAGGCGGCGGAGTTCGCCAGCGAGCGAGCGCGCA 120
290 AAGGCGGCAACGCGGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 343
121 AAGGCGGCAACGCGGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
344 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 397
181 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
398 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 457
241 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
458 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 517
301 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
518 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577
361 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
578 GCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 637
421 GCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
638 ACATGTGACCGGAGCTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 697
481 ACCTGTGACCGGAGCTGTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
698 CCGTGAAGCTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 757

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1e-165; Indels 18; Gaps 3;
Matches 2156; Conservative 0; Mismatches 138;

170 GCGGCAAGAGGCGCCACAGATGAAAGACTGACCCAGCGCCAGGCGCAACTTCTTCGCG 229
1 GGGGCGGGAAGGACCAATATGAAGATTGCACTGAGAGACAGGTATTTCTTCGCG 60
230 AGGACTGGGCTTCCCCCAAGGCGAGGCGCGAGTTTCCCAAGAGCAAAACCGCGCA 289
61 AGGACTGGGCTTCTTCTGCAAGGCGAGGCGCGAGTTTCAAGCGAGAGACCGCGCA 120
290 AAGAGCCCAACAGCCGCGAGCTGCAAGTCCGCGCG-----ACAACCCCGCAGAGAG 343
121 AAGAGCCCAACAGCCGCGAGCTGCAAGTCCGCGCGCGAGAAACAACCTGAGAGAG 180
344 CCGGCGCGAGGCGCCAGGCGACCTG-----AATTCCCGAGATCACTCTGTGACG 397
181 CCGGCGCGACCGCGAGGCGACCTGAGCTTCAACTTCCCGAGATCACTCTGTGACG 240
398 GCCCCTGTGAGCATCAAGTGGGCGGCAAGTCAAGAGGCGCTGTGACACCGGCG 457
241 GCCCCTGTGAGCATCAAGTGGGCGGCAAGTCAAGAGGCGCTGTGACACCGGCG 300
458 CCGAGCAACCTGTGTGAGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATG 517
301 CCGAGCAACCTGTGTGAGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATG 360
518 GCGGCACTGGGCGCTTCACTCAAGGTGCGCAAGTACAGACAGATCTGTATCGAGATGCG 577
361 GCGGCACTGGGCGCTTCACTCAAGGTGCGGAGTACAGACAGATCTCCGTGAGATGCG 420
578 GCAAGAGGCGCATGTGCGACGTGTGATCGGCCCAACCCCGTGAACATCATCGCGCG 637
421 GCCACAGGCGCATGTGCGACGTGTGATCGGCCCGCAACCCCGTGAACATCATCGCGCG 480
638 ACATGTGACCCAGCTGGGCTGCACTCTGAATTTCCCATCAGCCCCCATTCGAGACGTGC 697
481 ACCTGTGACCCAGATCGGCTGCACTCTGAATTTCCCATCAGCCCCCATTCGAGACGTGC 540
698 CCGTGAAGCTGAAGCCCGGATGGAAGGCGCCCAAGGTGAAGAGGCGCCCTGACCGAG 757
541 CCGTGAAGCTGAAGCCCGGAGTGAAGGCGCCCAAGGTGAAGAGGCGCCCTGACCGAG 600
758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGGAAGAGAGGCGCAAGATCA 817
601 AGAAGATCAAGGCGCTGTGTGAGATCTGACCGAGATGGAAGAGAGGCGCAAGATCA 660
818 AGATCGGCGCCGAGAACCCCTTCAACAACCCCGTGTTCGCCATCAAGAAAGAGACGA 877
661 AGATCGGCGCCGAGAACCCCTTCAACAACCCCGTGTTCGCCATCAAGAAAGAGACGA 720
878 CCAAGTGGGCGCAAGCTGTGTGAGATTTCCGAGAGTGAACAAGGCGCAACCGAGATTCTGG 937
721 CCAAGTGGGCGCAAGCTGTGTGAGATTTCCGAGAGTGAACAAGGCGCAACCGAGATTCTGG 780
938 AGGTGAGCTGGGCACTCCCGACCCCGCGGCTGTGAAGAAAGAGAGAGGCTGACGCTGC 997
781 AGGTGAGCTGGGCACTCCCGACCCCGCGGCTGTGAAGAAAGAGAGAGGCTGACGCTGC 840
998 TGAAGCTGGGCGAGCGCTTACTGAGCGTGCCTGTGACGAGAGATTCCCGAGATCAACCG 1057
841 TGAAGCTGGGCGAGCGCTTACTGAGCGTGCCTGTGACGAGAGATTCCCGAGATCAACCG 900
1058 CCTTCAACATCCCGAGATCAACAAGAGACCCCGGCAATCCGCTACAGTAAAGGTG 1117
901 CCTTCAACATCCCGAGATCAACAAGAGACCCCGGCAATCCGCTACAGTAAAGGTG 960
1118 TGGCCAGAGGCTGAAGAGGCGCCAGCATCTTTCAGAGACAGATGACCAAGATCTGG 1177
961 TGGCCAGAGGCTGAAGAGGCGCCAGCATCTTTCAGAGACAGATGACCAAGATCTGG 1020
1178 AGCCCTTCGCGCGCGGCAACCCGAGATGTGATCTACAGTACATGAGACGACTGTAGC 1237

1021 AGCCCTTCGAGAGAGAACCCCGACATCTGTATTTACCA-----GGCCCCCTGTAGC 1074
1238 TGGGAGGAGCCTTGAGATCTGGCGAGCACCGCGCAAGATCGAGAGCTGCGCAAGACC 1297
1075 TGGGAGGAGCCTTGAGATCTGGCGAGCACCGGCAACAAGATCGAGAGCTGCGCGACACC 1134
1298 TGTCTGCTGGGCGCTTCAACACCCCGGCAAGAAACCAAGAAAGAGGCGCCCTTCTGT 1357
1135 TGTCTGCTGGGCGCTTCAACACCCCGGCAAGAAACCAAGAAAGAGGCGCCCTTCTGT 1194
1358 GGAATGGGCTACAGAGCTGACCCCGCAAGTGAACGCTGACGCGCATTCAGAGCTGCGAG 1417
1195 GGAATGGGCTACAGAGCTGACCCCGCAAGTGAACGCTGACGCGCATTCAGAGCTGCGAG 1254
1418 AGGAGAGCTGAGCCGTGAACGACATCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCC 1477
1255 AGGAGAGCTGAGCCGTGAACGACATCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCC 1314
1478 AGATTTACCCCGGCAATCAAGTGGCGCAGCTGTGCAAGCTGTGGCGCGCGCAAGGCGCC 1537
1315 AGATTTACCGCGCAATCAAGTGAAGAGCTGTGCAAGCTGTGGCGCGCGCAAGGCGCC 1374
1538 TGAACGACATGTGCGCCCTGACCGAGAGGCGGAGCTGAGGCTGGCGCGAGAACCGGAG 1597
1375 TGAACGAGGTATTCCTTGAACGAGAGGCGGAGCTGAGGCTGGCGCGAGAACCGGAG 1434
1435 TCTTGAAGAGGCGCGTGAACGAGAGTGTACTAGACCCCGAGAAAGGCGTGTGGCGGAGA 1494
1658 TCCAGAGAGGCGCGACGACAGTGAACCTTACAGATCTACAGAGAGGCTTCAAGAAC 1717
1495 TCCAGAGAGGCGCGAGGCGAGTGAACCTTACAGATCTACAGAGAGGCTTCAAGAAC 1554
1718 TGAAGAGGCGGAGTACGCAAGATGGCGCACCGGCGCAACCAAGAGCTGAAGAGCTGA 1777
1555 TGAAGAGGCGGAGTACGCGAGAGCGGCGCGCGCGCAACCAAGAGCTGAAGAGCTGA 1614
1778 CCGAGGCGCTGAGAGAGATCGCATGAGAGCATGTGATCTTGGGCGAAGACCCCGAGT 1837
1615 CCGAGGCGCTGAGAGAGATCGAGAGCATGTGATCTTGGGCGAAGATCCCGAGT 1674
1838 TCGCGCTGCGCATCAAGAGAGAGACCTTGGAGACCTGTGTGACCGACTTCTGCGAGCGCA 1897
1675 TCAAGCTGCGCATCAAGAGAGAGACCTTGGAGAGCGCTGTGTGATGTGAGTCTGCGAGCGCA 1734
1898 CCTGATCCCCAGATGGGAGTTCGTGAACAACCCCGCTGTGTGAAGCTGTGTACCAAC 1957
1735 CCTGATCCCCAGATGGGAGTTCGTGAACAACCCCGCTGTGTGAAGCTGTGTACCAAC 1794
1958 TGAAGAGAGGCGCATCATCTGCGCGCGAGACCTTCTTACGTGAGACGCGCGCAACCGCG 2017
1795 TGAAGAGAGGCGCATCTGTGGCGCGAGACCTTCTTACGTGAGACGCGCGCAACCGCG 1854
2018 AGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGCAAGATGTGAGGC 2077
1855 AGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGCAAGATGTGAGGC 1914
2078 TGAACGAGACCAACCAAGAGAGACCTGAGTCAAGAGGCGCGCTGCGAGAGCA 2137
1915 TGGCGGACCAACCAAGAGAGACCTGAGTCAAGAGGCGCGCTGCGAGAGCA 1974
2138 GCGGAGAGGCTGAACATCTGTGAACGAGACGAGCATCGCTGTGGCATCATCAGGCGCC 2197
1975 GCGGAGGCTGAAGATCTGTGAACGAGACGAGCATCGCTGTGGCATCATCAGGCGCC 2034
2198 AGCCCGACAAGAGCGAGAGGCTGTGAGCAAGATCAATCAAGAGGCTGTATCAAGAGG 2257
2035 AGCCCGACAAGAGCGAGAGGCTGTGAGCAAGATCAATCAAGAGGCTGTATCAAGAGG 2094
2258 AGAAGGTATCTGTAGCTGGGCGCGCGCAAGAGGCAATCGCGCGCAACGAGCATGCG 2317

DB 1555 TGAAGACCCGGCAGATCGCCCGCATCGCGCGCCCAACACAGCTGAAGCACTGA 1614
QY 1778 CCGAGGCCGTGCAAGAGATCGCCATGAGAGCATCTGTATCTTGGGCAAGACCCCAAGT 1837
DB 1615 CCGAGGCCGTGCAAGAGATGAGACCGAGAGCATCTGTATCTTGGGCAAGATCCCAAGT 1674
QY 1838 TCGCGCTGCTCCATCCAGAGAGAGACTGGGAGACTGTGTGACCCACTACTGGCAGGCCA 1897
DB 1675 TCAGCTGCCCATCCAGAGAGAGACTGGGAGGCTGTGTGATGAGTACTGGCAGGCCA 1734
QY 1898 CCTGATCCCGAGTGGAGATTGTGAACACCCCGCCCTGTGTGAGCTGTGTACAGC 1957
DB 1735 CTTGATCCCGAGTGGAGATTGTGAACACCCCGCCCTGTGTGAGCTGTGTACAGC 1794
QY 1958 TGGAGAGAGAGCCCATCATCGCGCGCCAGACCTTCTACGTGACCGCGCCGCCACCGCG 2017
DB 1795 TGGAGAGAGAGGCCCATCGTGGCGCCGAGACCTTCTACGTGACCGCGCCGCCACCGCG 1854
QY 2018 AGACCAAGATCGGCAAGCGCGGCTACGTGACCGACCGGGCGCGCAAGAGATGTGAGCC 2077
DB 1855 AGACCAAGCTGGCAAGCGCGGCTACGTGACCGACCGGGCGCGCAAGAGTGTGAGCA 1914
QY 2078 TGAACGAGACCAACCAAGAGAGACCGAGCTGAGGCAATCCAGCTGGCCCTGCAAGACA 2137
DB 1915 TCGCCGACACACCAACCAAGAGACCGAGCTGAGGCAATCCAGCTGGCCCTGCAAGACA 1974
QY 2138 GCGGCAAGAGAGTGAACATCTGTGACCGACAGCCAGTACGCTGGGCAATCATCAGGCC 2197
DB 1975 GCGGCTGAGAGTGAACATCTGTGACCGACAGCCAGTACGCTGGGCAATCATCAGGCC 2034
QY 2198 AGCCGACCAAGAGCGAGAGCTGTGAACTAGATCATCGACAGCTGATCAAGAGG 2257
DB 2035 AGCCGACCAAGAGCGAGAGCTGTGAACTAGATCATCGACAGCTGATCAAGAGG 2094
QY 2258 AGAAGGTGTACTGAGCTGGGTGCCCGCCCAAGAGGCAATCGGGGCAACGAGCAGATCG 2317
DB 2095 AGAAGGTGTACTGAGCTGGGTGCCCGCCCAAGAGGCAATCGGGGCAACGAGCAGGTGG 2154
QY 2318 ACAAGCTGTGAGCAAGGCAATCGCAAGGTGCTGTCTGTGACGGCATCGATGGCGCA 2377
DB 2155 ACAAGCTGTGAGCGCGCGCATCCGCAAGGTGCTGTCTGTGACGGCATCGATGGCGCA 2214
QY 2378 TCGTGAATCTACAGTACATGAGACGACTGTACGTGGGCAAGCGGCGGCTTAGATCGATT 2437
DB 2215 TCGTGAATCTACAGTACATGAGACGACTGTACGTGGGCAAGCGGCGGCTTAGATCGATT 2274
QY 2438 AAAAGCTTCCGGGGCTAGACCGGTGAATTTC 2469
DB 2275 AAAAGCTTCCGGGGCTAGACCGGTGAATTTC 2306

Search completed: December 30, 2005, 19:16:16
Job time : 11995.5 secs

This Page Blank (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1299.83 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469
Sequence: 1 gtcgacgccaccatgcccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21:*

- 1: geneseqn1980a:*
- 2: geneseqn1990a:*
- 3: geneseqn2000a:*
- 4: geneseqn2001a:*
- 5: geneseqn2001b:*
- 6: geneseqn2002a:*
- 7: geneseqn2002b:*
- 8: geneseqn2003a:*
- 9: geneseqn2003b:*
- 10: geneseqn2003c:*
- 11: geneseqn2003d:*
- 12: geneseqn2004a:*
- 13: geneseqn2004b:*
- 14: geneseqn2005a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6	ABL39959
2	2469	100.0	2469	12	ADM73764
3	2457	99.5	2457	8	ACA03548
4	2457	99.5	2457	10	ADCl3266
5	2442.2	98.9	2463	6	ABL39960
6	2442.2	98.9	2463	12	ADM73765
7	2434.6	98.6	2457	8	ACA03547
8	2434.6	98.6	2457	10	ADCl3265
9	2415.4	97.8	2457	6	ABL39961
10	2415.4	97.8	2457	12	ADM73766
11	2401.8	97.3	2445	8	ACA03546
12	2401.8	97.3	2445	10	ADCl3264
13	2394.8	97.0	3930	10	ADCl3230
14	2393.2	96.9	3930	10	ADCl3231
15	2393.2	96.9	3930	10	ADCl3232
16	2393.2	96.9	5184	8	ACA03591
17	2393.2	96.9	5184	10	ADCl3279
18	2362.8	95.7	3531	10	ADCl3234
19	2361.2	95.6	3537	10	ADCl3236

20	2360.2	95.6	5145	8	ACA03521	ACA03521 Synthetic
21	2360.2	95.6	5145	10	ADCl3233	ADCl3233 DNA of HI
22	2350.2	95.2	3538	10	ADCl3235	ADCl3235 DNA of HI
23	2349.4	95.2	3624	8	ACA03550	ACA03550 Synthetic
24	2349.4	95.2	3624	10	ADCl3268	ADCl3268 DNA of HI
25	2301.8	93.2	3607	8	ACA03551	ACA03551 Synthetic
26	2301.8	93.2	3607	10	ADCl3269	ADCl3269 DNA of HI
27	2283.6	92.5	3597	10	ACA03549	ACA03549 Synthetic
28	2283.6	92.5	3597	10	ADCl3267	ADCl3267 DNA of HI
29	2149.6	87.1	2472	8	ACA03543	ACA03543 Synthetic
30	2149.6	87.1	2472	8	ACC78507	ACC78507 HIV p2Pol
31	2121.2	85.9	2466	8	ACA03542	ACA03542 Synthetic
32	2121.2	85.9	2466	8	ACC78506	ACC78506 HIV p2Pol
33	2094.4	84.8	2460	8	ACA03541	ACA03541 Synthetic
34	2094.4	84.8	2460	8	ACC78505	ACC78505 HIV p2Pol
35	2093.4	84.8	3564	8	ACC78488	ACC78488 HIV GagPo
36	2093.4	84.8	3564	8	ACC78489	ACC78489 HIV GagPo
37	2092.8	84.8	4716	8	ACA03522	ACA03522 Synthetic
38	2092.8	84.8	4716	10	ADCl3238	ADCl3238 DNA of HI
39	2089	84.6	3999	8	ACC78484	ACC78484 HIV GagCo
40	2087.4	84.5	3999	8	ACC78485	ACC78485 HIV GagCo
41	2087.4	84.5	3999	8	ACC78486	ACC78486 HIV GagCo
42	2087.4	84.5	5283	8	ACA03584	ACA03584 Synthetic
43	2087.4	84.5	5283	8	ACC78529	ACC78529 HIV TatRe
44	2087.2	84.5	4713	8	ACA03592	ACA03592 Synthetic
45	2087.2	84.5	4713	10	ADCl3280	ADCl3280 DNA of HI

ALIGNMENTS

RESULT 1	ABL39959	standard; DNA; 2469 BP.
ID	ABL39959	standard; DNA; 2469 BP.
XX	ABL39959;	
AC	15-MAY-2002 (first entry)	
DT	Synthetic construct PR975(+) SEQ ID NO:30.	
XX	Human immunodeficiency virus type C; antigenic HIV type C protein;	
KW	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;	
KW	immunostimulant; gene therapy; gene; ds.	
OS	Human immunodeficiency virus; type C.	
XX	Synthetic.	
XX	WO200204493-A2.	
XX	17-JAN-2002.	
PD	05-JUL-2001; 2001WO-US021241.	
PF	05-JUL-2000; 2000US-00610313.	
PR	(CHIR) CHIRON CORP.	
XX	(UYST-) UNIV STELLENBOSCH.	
PA	Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	
XX	WPI; 2002-154920/20.	
DR	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	
PT	applications including DNA immunization or generation of packaging cell	
PT	lines, particularly in gene therapy.	
XX	Claim 1; Fig 8; 233p; English.	
PS	The present invention describes expression cassettes comprising a	
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	
CC	type C polypeptides. The expression cassettes comprise any of the HIV	
CC	type C sequences encoding gag, pol, vif, vpr, tat, rev, env or nef	

CC (1). (1) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX

SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 8.3e-298;

Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGACGACCACTGACGAGGCGCATGACGAGCCACGAGGCGCCAACTCCGATGACG 60
DB 1 GTGACGACCACTGACGAGGCGCATGACGAGCCACGAGGCGCCAACTCCGATGACG 60
QY 61 CGCAGCACTTCAAGGAGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGAGC 120
DB 61 CGCAGCACTTCAAGGAGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGAGC 120
QY 121 CACATGCGCCGCACTGCGCGCCGCCCGGCAAGAGGCTGCTGGAAGTGCAGAGAG 180
DB 121 CACATGCGCCGCACTGCGCGCCGCCCGGCAAGAGGCTGCTGGAAGTGCAGAGAG 180
QY 181 GGCACACAGATGAAGGACTGCAACGAGGCGCAGGCGCAACTTCTTCCGAGAGGACCTGAGC 240
DB 181 GGCACACAGATGAAGGACTGCAACGAGGCGCAGGCGCAACTTCTTCCGAGAGGACCTGAGC 240
QY 241 TTCCCCCAAGGAGGAGGCGGAGGTTCCCGACGAGCAGAGACCGCGCCAAAGCCCGACC 300
DB 241 TTCCCCCAAGGAGGAGGCGGAGGTTCCCGACGAGCAGAGACCGCGCCAAAGCCCGACC 300
QY 301 AGCCCGAGAGTCAAGTGTGCGCGGCAACACCCCGAGAGAGGCGCGCGCGGCGGCGAG 360
DB 301 AGCCCGAGAGTCAAGTGTGCGCGGCAACACCCCGAGAGAGGCGCGCGCGGCGGCGAG 360
QY 361 GGCACCGTGAATCTTCCCGAGATCACTGTGAGAGCGCCCTCGTGAAGCATCAAGGTG 420
DB 361 GGCACCGTGAATCTTCCCGAGATCACTGTGAGAGCGCCCTCGTGAAGCATCAAGGTG 420
QY 421 GCGCGGCAAGATCAAGAGAGGCGCTGTGAGACCGCGCGCAGACGACGCTGTGAGAG 480
DB 421 GCGCGGCAAGATCAAGAGAGGCGCTGTGAGACCGCGCGCAGACGACGCTGTGAGAG 480
QY 481 ATGAGGCTTCCCGGCAAGTGAAGCCCAAGATGATGCGCGCATCGCGCTTCAATCAAG 540
DB 481 ATGAGGCTTCCCGGCAAGTGAAGCCCAAGATGATGCGCGCATCGCGCTTCAATCAAG 540
QY 541 GTGCGGCAATGAAGACCAAGATCTGATGAGATCTGCGGCAAGAGGCGCATTCGGCACCGTG 600
DB 541 GTGCGGCAATGAAGACCAAGATCTGATGAGATCTGCGGCAAGAGGCGCATTCGGCACCGTG 600
QY 601 CTGATCGAGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCGAGTGGCTGC 660
DB 601 CTGATCGAGCCCAACCCCGGTGAACATCATCGCGCGCAACATGCTGACCCGAGTGGCTGC 660
QY 661 ACCCTGAATCTTCCCATCAAGCCCATCGAGACCGTGTCCGTGAAGCTGAAGCCCGGCAAG 720
DB 661 ACCCTGAATCTTCCCATCAAGCCCATCGAGACCGTGTCCGTGAAGCTGAAGCCCGGCAAG 720
QY 721 GACGAGCCCAAGGATGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGAACCGGC 780
DB 721 GACGAGCCCAAGGATGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCGCTGAACCGGC 780
QY 781 ATCTGGAAGAGATGAGAGAGAGGCAAGATCAACAGATCGCGCCCGAGAACCCCTAC 840
DB 781 ATCTGGAAGAGATGAGAGAGAGGCAAGATCAACAGATCGCGCCCGAGAACCCCTAC 840
QY 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGTGAC 900
DB 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGTGAC 900
```

```
QY 901 TTCCCGAGCTGAACAAAGGACCCAGAGACTTCTGAGAGTGCAGCTGGGACATCCCCAC 960
DB 901 TTCCCGAGCTGAACAAAGGACCCAGAGACTTCTGAGAGTGCAGCTGGGACATCCCCAC 960
QY 961 CCCGCGGCGTGAAGAGAGAGAGCGTGAACCGTGTCTGAGAGTGGCGGACGCTTACTTC 1020
DB 961 CCCGCGGCGTGAAGAGAGAGAGCGTGAACCGTGTCTGAGAGTGGCGGACGCTTACTTC 1020
QY 1021 AGCGTGGCCCTGGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACATCCCGAGCATCAAC 1080
DB 1021 AGCGTGGCCCTGGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACATCCCGAGCATCAAC 1080
QY 1081 AACGAGACCCCGGAGATCGCTTACAGTACAAAGTGTGCGCCGAGGCTGGAAGAGGAGC 1140
DB 1081 AACGAGACCCCGGAGATCGCTTACAGTACAAAGTGTGCGCCGAGGCTGGAAGAGGAGC 1140
QY 1141 CCCAGCATCTTCCAGAGAGACATGAACAAAGATCTGAGAGCCCTTCCGCGCCGCAACCCC 1200
DB 1141 CCCAGCATCTTCCAGAGAGACATGAACAAAGATCTGAGAGCCCTTCCGCGCCGCAACCCC 1200
QY 1201 GAGATCGTGAATTAACAGTACATGAGACGACTGTACCTGTGGGAGGAGCCTTGAAGTCCGC 1260
DB 1201 GAGATCGTGAATTAACAGTACATGAGACGACTGTACCTGTGGGAGGAGCCTTGAAGTCCGC 1260
QY 1261 CAGCACCGGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGCTGGGAGCTTCAACACC 1320
DB 1261 CAGCACCGGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGCTGGGAGCTTCAACACC 1320
QY 1321 CCCGACAAAGAGACACAGAGAGGCGCCCTTCTGTGATGAGGCTTACAGCTGACACCC 1380
DB 1321 CCCGACAAAGAGACACAGAGAGGCGCCCTTCTGTGATGAGGCTTACAGCTGACACCC 1380
QY 1381 GACAAAGTGAACCGTGAACCGCATTCGAGCTGCCGAGAGAGAGAGCTGACCGTGAACGAC 1440
DB 1381 GACAAAGTGAACCGTGAACCGCATTCGAGCTGCCGAGAGAGAGAGCTGACCGTGAACGAC 1440
QY 1441 ATTCAGAAAGCTGTGGGCAAGCTGAAGTGAAGCTGAGCAACAGATCTTACCCCGGACATCAAGGTG 1500
DB 1441 ATTCAGAAAGCTGTGGGCAAGCTGAAGTGAAGCTGAGCAACAGATCTTACCCCGGACATCAAGGTG 1500
QY 1501 CGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCTTGAACGACATCTGTGCGCTTGAAC 1560
DB 1501 CGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCGCTTGAACGACATCTGTGCGCTTGAAC 1560
QY 1561 GAGGAGGCGGAGCTGAGAGCTGCGCCAGAACCGCGAGATCTCTGCGGAGCGCTTGAACGAGC 1620
DB 1561 GAGGAGGCGGAGCTGAGAGCTGCGCCAGAACCGCGAGATCTCTGCGGAGCGCTTGAACGAGC 1620
QY 1621 GTGTACTAGACCCGAGCAAGGACCTGTGCGGAGATTCAGAGACGAGGCGCACGACGAG 1680
DB 1621 GTGTACTAGACCCGAGCAAGGACCTGTGCGGAGATTCAGAGACGAGGCGCACGACGAG 1680
QY 1681 TGAACCTTACAGATCTTACAGAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCAAG 1740
DB 1681 TGAACCTTACAGATCTTACAGAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCAAG 1740
QY 1741 ATGCGCACGCGCCACACCAAGAGCTGAAGAGAGCTGACCGAGGCGGTGAGAGAGATCGGC 1800
DB 1741 ATGCGCACGCGCCACACCAAGAGAGCTGAAGAGAGCTGACCGAGGCGGTGAGAGAGATCGGC 1800
QY 1801 ATGAGAGCATGTGATCTTGGGCAAGACCCCGCAAGTTCGCGCTGCGCATCCAGAGAGAG 1860
DB 1801 ATGAGAGCATGTGATCTTGGGCAAGAGACCCCGCAAGTTCGCGCTGCGCATCCAGAGAGAG 1860
QY 1861 ACCTGGAGAGCTGTGTGAACCGACTTACGAGGCGACCTGTGATCCCGAGTGGAGTTTC 1920
DB 1861 ACCTGGAGAGCTGTGTGAACCGACTTACGAGGCGACCTGTGATCCCGAGTGGAGTTTC 1920
QY 1921 GTGAACACCCCGCTGTGTGAAGCTGTGTACGAGCTGAGAGAGAGGCGCATCATTCGCG 1980
DB 1921 GTGAACACCCCGCTGTGTGAAGCTGTGTACGAGCTGAGAGAGAGGCGCATCATTCGCG 1980
```

QY 1981 GCCGAGACCTTTCTACGTGACCGGCGCCGCAACCGGAGACCAAGATCCGCAAGGCGCGC 2040
DB 1981 GCCGAGACCTTTCTACGTGACCGGCGCCGCAACCGGAGACCAAGATCCGCAAGGCGCGC 2040
QY 2041 TACGTACCGACCGGCGCGCCGCAAGATCGTGAAGCTGACCGAGACCAACCAAGG 2100
DB 2041 TACGTACCGACCGGCGCGCCGCAAGATCGTGAAGCTGACCGAGACCAACCAAGG 2100
QY 2101 ACCGAGCTGACGCGCATTCACGTGCGCTTGACAGACGCGGACGAGGTGAACATCGTG 2160
DB 2101 ACCGAGCTGACGCGCATTCACGTGCGCTTGACAGACGCGGACGAGGTGAACATCGTG 2160
QY 2161 ACCGAGCTGACGCGCATTCACGTGCGCTTGACAGACGCGGACGAGGTGAACATCGTG 2220
DB 2161 ACCGAGCTGACGCGCATTCACGTGCGCTTGACAGACGCGGACGAGGTGAACATCGTG 2220
QY 2221 CTGCTGAACCAAGATTCATCGACGAGCTGATCAAGAGAGAGGTGTAACCTGAGCTGAGTG 2280
DB 2221 CTGCTGAACCAAGATTCATCGACGAGCTGATCAAGAGAGAGGTGTAACCTGAGCTGAGTG 2280
QY 2281 CCCGCCACCAAGGAGCATTCGCGCGCAACGAGCAGATGACAGAGCTGTGAGCAAGGCGATC 2340
DB 2281 CCCGCCACCAAGGAGCATTCGCGCGCAACGAGCAGATGACAGAGCTGTGAGCAAGGCGATC 2340
QY 2341 CGCAAGGTGCTGTTCTCTGACGCGCATTCGATGCGCGCATCTGATCTAACGATCATGAGAC 2400
DB 2341 CGCAAGGTGCTGTTCTCTGACGCGCATTCGATGCGCGCATCTGATCTAACGATCATGAGAC 2400
QY 2401 GACCTGTAGCTGAGGACGCGCGCGCTTGAAGATGATTAAAGGCTCCCGGCGCTAGACCC 2460
DB 2401 GACCTGTAGCTGAGGACGCGCGCGCTTGAAGATGATTAAAGGCTCCCGGCGCTAGACCC 2460
QY 2461 GGTGAATTC 2469
DB 2461 GGTGAATTC 2469

RESULT 2

ADM73764

ID ADM73764 standard; DNA; 2469 BP.

AC ADM73764;

XX 03-JUN-2004 (first entry)

XX HIV-1 polynucleotide #7.

XX HIV-1; gene; de; HIV pol; immune response; DNA immunisation;

KW HIV type C protein; immunostimulant.

XX Human immunodeficiency virus 1.

XX US2003223961-A1.

XX 04-DEC-2003:

XX 05-JUL-2001; 2001US-00899575.

XX 05-JUL-2000; 2000US-00610313.

XX (MEGE/) MEGEDE J Z.

XX (BARN/) BARNETT S W.

XX (ENGEL/) ENGELBRECHT S.

XX (RENS/) RENSBURG B J V.

XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;

XX WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an

XX HIV pol polypeptide, useful in eliciting an immune response, in DNA

XX immunization, generating of packaging cell lines or in producing HIV type

XX C proteins.

XX PS Claim 1: SEQ ID NO 30; 160bp; English.
XX CC The invention relates to an expression cassette comprising a
XX CC polynucleotide sequence encoding an HIV pol polypeptide. The invention
XX CC also relates to a recombinant expression system for use in a host cell
XX CC comprising an expression cassette, where the polynucleotide sequence
XX CC further comprises control elements capable of driving expression in the
XX CC selected host cell, a cell comprising an expression cassette where the
XX CC polynucleotide sequence further comprises control elements compatible
XX CC with the expression in the cell and a composition for generating an
XX CC immunological response, comprising an expression cassette. The expression
XX CC cassette and the method of the invention are useful in eliciting an
XX CC immune response, in DNA immunisation, in generation of packaging cell
XX CC lines and in producing HIV type C proteins. This sequence represents an
XX CC HIV-1 polynucleotide of the invention.
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 12; Length 2469;
Best Local Similarity 100.0%; Pred. No. 8.3e-298;
Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGACGCGCACCATGCGCGGCGCATGAGCGGCGGACCGGCGGCAATCTGATGACG 60
DB 1 GTGACGCGCACCATGCGCGGCGCATGAGCGGCGGACCGGCGGCAATCTGATGACG 60
QY 61 CGGAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGCG 120
DB 61 CGGAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGGCAAGAGGCG 120
QY 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 GGCACACAGATGAAGGATGTCACCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
DB 181 GGCACACAGATGAAGGATGTCACCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 241 TTCGCCACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 241 TTCGCCACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 301 AGCGCGAGCTGCGAGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
DB 301 AGCGCGAGCTGCGAGGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 361 GGCACCTTGAACCTTCCCGGAGATCACTGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 361 GGCACCTTGAACCTTCCCGGAGATCACTGCTGCAAGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 421 GCGCGCGAGATCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 421 GCGCGCGAGATCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 ATGAGCTTGGCGCGGCAAGTGAAGCGGCAAGATATGCGCGCGCGCGCGCGCGCGCGCGCG 540
DB 481 ATGAGCTTGGCGCGGCAAGTGAAGCGGCAAGATATGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 GTGCGCGAGTGAAGCAAGATCTGATCGAGATCTGCGGCGGCAAGAGGCGCATCGGCGCGTG 600
DB 541 GTGCGCGAGTGAAGCAAGATCTGATCGAGATCTGCGGCGGCAAGAGGCGCATCGGCGCGTG 600
QY 601 CTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGAATCTTCCCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 ACCCTGAATCTTCCCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 GACGCGCGCGAGGATGAAGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 GACGCGCGCGAGGATGAAGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

Db 721 GACGGCCCCAAGTGAAGCATGTGGCCCTTGAACGAGAGAAATCAAGGCCCTTGAACCGCC 780
Qy 781 ATCTCGAGAGAGATGGAAGAGAGGCGCAAGATCAACAAATGCGGCCCGAGAACCCCTTAC 840
Db 781 ATCTCGAGAGAGATGGAAGAGAGGCGCAAGATCAACAAATGCGGCCCGAGAACCCCTTAC 840
Qy 841 AACACCCCCGTGTGCGCATCAAGAGAGAGACAGCAAGTGGCGCAAGCTGTGTGAC 900
Db 841 AACACCCCCGTGTGCGCATCAAGAGAGAGACAGCAAGTGGCGCAAGCTGTGTGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCAACCGAGACTTCTGGAAGTGCAGCTGGGCAATCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCAACCGAGACTTCTGGAAGTGCAGCTGGGCAATCCCCAC 960
Qy 961 CCCGCGCGCTGAAGAGAGAGAGACCGTGAACCGTGTGAGAGTGGGCGAACCCCTTACTTC 1020
Db 961 CCCGCGCGCTGAAGAGAGAGAGACCGTGAACCGTGTGAGAGTGGGCGAACCCCTTACTTC 1020
Qy 1021 AACGTCGCCCTGGAACGAGACTTCCGCAAGTACAACGCGCTTGAACATCCCGACAGCATCAAC 1080
Db 1021 AACGTCGCCCTGGAACGAGACTTCCGCAAGTACAACGCGCTTGAACATCCCGACAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCAATCCGCTTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCAATCCGCTTACCAAGTACAAAGTGTGCTGCCAGGGCTGGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGAGAGCATGACCAAGATCTGTGAAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGAGAGCATGACCAAGATCTGTGAAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATGCTGATCTACCAAGTACATGAGACGCTGTGAGTGGGCGACGACCTTGAAGATCGAC 1260
Db 1201 GAGATGCTGATCTACCAAGTACATGAGACGCTGTGAGTGGGCGACGACCTTGAAGATCGAC 1260
Qy 1261 GAGCACCAGCGGCAAGATCGAGAGAGCTGCGCAAGCACTGCTGCGCTTCAACCAAC 1320
Db 1261 GAGCACCAGCGGCAAGATCGAGAGAGCTGCGCAAGCACTGCTGCGCTTCAACCAAC 1320
Qy 1321 CCCGCAAAAGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 CCCGCAAAAGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy 1381 GACAAAGTGAACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 GACAAAGTGAACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1441 ATTCAGAAAGCTGTGTGGGCAAGCTGAACCTGAGGCGAGCAGATCTAACCCCGCATCAAGGTG 1500
Db 1441 ATTCAGAAAGCTGTGTGGGCAAGCTGAACCTGAGGCGAGCAGATCTAACCCCGCATCAAGGTG 1500
Qy 1501 CGCCAGCTGTGCAAGCTGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1560
Db 1501 CGCCAGCTGTGCAAGCTGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1560
Qy 1561 GAGAGAGGCGGAGCTGAGAGCTGAGGCGGAGAACCGGAGATCTGTGCGAGAGCGCGTGAACGCGC 1620
Db 1561 GAGAGAGGCGGAGCTGAGAGCTGAGGCGGAGAACCGGAGATCTGTGCGAGAGCGCGTGAACGCGC 1620
Qy 1621 GTGTACTAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1621 GTGTACTAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1681 TGGACCTTACCAATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1681 TGGACCTTACCAATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy 1741 ATGCGCACCGGCGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1741 ATGCGCACCGGCGCACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1801 ATGAGAGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Db 1801 ATGAGAGAGATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860

Qy 1861 ACTTGGAGAGACTGTGTGACCGAATCTATCTGGAGAGGCACTGTGATTCGCCAGTGGAGATTTC 1920
Db 1861 ACTTGGAGAGACTGTGTGACCGAATCTATCTGGAGAGGCACTGTGATTCGCCAGTGGAGATTTC 1920
Qy 1921 GTGAACACCCCGCCCTGTGTGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Db 1921 GTGAACACCCCGCCCTGTGTGAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Qy 1981 GCCAGACCTTCTTACGTGAGCGGCGCGCAACCGCGAGAGCCAAATGTGGCAATGGCCCGC 2040
Db 1981 GCCAGACCTTCTTACGTGAGCGGCGCGCAACCGCGAGAGCCAAATGTGGCAATGGCCCGC 2040
Qy 2041 TACGTGAACCGAGCCGGGCGGCGAGAGAGATCGTGAAGCTGAGCCGAGACCAACCAAGAG 2100
Db 2041 TACGTGAACCGAGCCGGGCGGCGAGAGAGATCGTGAAGCTGAGCCGAGACCAACCAAGAG 2100
Qy 2101 ACCGAGCTGAGAGGCAATCGAGCTGCGCCCTGACAGACAGCGGAGGAGTGAACATCGTG 2160
Db 2101 ACCGAGCTGAGAGGCAATCGAGCTGCGCCCTGACAGACAGCGGAGGAGTGAACATCGTG 2160
Qy 2161 ACCGAGCGCAGTACGCGCTTGGGCAATCTCAAGGCTCCAGCCGACCAAGAGCGAGAGCGAG 2220
Db 2161 ACCGAGCGCAGTACGCGCTTGGGCAATCTCAAGGCTCCAGCCGACCAAGAGCGAGAGCGAG 2220
Qy 2221 CTGTGGAACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 CTGTGGAACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy 2281 CCCGCCCAAGAGGAGATCGGCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 CCCGCCCAAGAGGAGATCGGCGGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2341 CGCAAGTGTGTCTTCTGAGCGGCAATCGATGCGGCAATCGATCTTACCAATGATGAC 2400
Db 2341 CGCAAGTGTGTCTTCTGAGCGGCAATCGATGCGGCAATCGATCTTACCAATGATGAC 2400
Qy 2401 GACCTGTACGTGGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2460
Db 2401 GACCTGTACGTGGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2460
Qy 2461 GGTGAATTC 2469
Db 2461 GGTGAATTC 2469

RESULT 3
ID ACA03548
ID ACA03548 standard; DNA; 2457 BP.
AC
ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
XX
DE Synthetic DNA encoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
OS Synthetic.
XX
PN MO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.

XX (CHTR) CHIRON CORP.
PA Zur Megele J, Barnett SW, Lian Y;
PI WPI; 2003-221602/21.
XX
XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
XX Example 1; Fig 36; 262bp; English.
XX
XX The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
XX Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
SQ
Query Match 99.5%; Score 2457; DB 8; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2.6e-296;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 GCCACCATGCGCGAGCCATGAGCCAGGCCACCAAGCCCACTCTGATGACGCGCAGC 66
DB 1 GCCACCATGCGCGAGCCATGAGCCAGGCCACCAAGCCCACTCTGATGAGCGCGCAGC 60
QY 67 AACTTCAAGGGGCCCAAGGCGCATCATCAAGTGTCTTCAATCGCGGCAAGAGGGCCCATC 126
DB 61 AACTTCAAGGGGCCCAAGGCGCATCATCAAGTGTCTTCAATCGCGGCAAGAGGGCCCATC 120
QY 127 GCCCGCACTGCGCGGCGCGCGCGCAAGAGGGGCTGTGAGAGTGGCGGCAAGAGGGCGC 186
DB 121 GCCCGCACTGCGCGGCGCGCGCGCAAGAGGGGCTGTGAGAGTGGCGGCAAGAGGGCGC 180
QY 187 CAGATGAAGACTGCAACCGAGCGCCAGGCCCACTTCTTCGCGAGGAGCCTGAGCTTCCCG 246
DB 181 CAGATGAAGACTGCAACCGAGCGCCAGGCCCACTTCTTCGCGAGGAGCCTGAGCTTCCCG 240
QY 247 CAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAGACCGCGCCAAAGCCCCCAGCCCGC 306
DB 241 CAGGGCAAGGCCCGCGAGTTCCTCCAGCGAGCAGACCGCGCCAAAGCCCCCAGCCCGC 300
QY 307 GAGCTGCAAGTGGCGGCGGCAACCCCCCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGC 366
DB 301 GAGCTGCAAGTGGCGGCGGCAACCCCCCGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGC 360
QY 367 CTGGAATTCCTCCCGAGATCACTCTGTGAGAGCGCCCTGCTGAGCATCAAGGTGGCGCGC 426
DB 361 CTGGAATTCCTCCCGAGATCACTCTGTGAGAGCGCCCTGCTGAGCATCAAGGTGGCGCGC 420
QY 427 CAGATCAAGAGAGCCCTGCTGTGACACCGGCGCGAGCAGCACCGTGTGTGAGAGATGAGC 486
DB 421 CAGATCAAGAGAGCCCTGCTGTGACACCGGCGCGAGCAGCACCGTGTGTGAGAGATGAGC 480
QY 487 CTGCGCGGCAAGTGGAGGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGC 546
DB 481 CTGCGCGGCAAGTGGAGGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGC 540
QY 547 CAGTACGACCAAGATCTGTGATGAGATCTGTGCGGCAAGAGGCGCATCGGCACTGTGATG 606
DB 541 CAGTACGACCAAGATCTGTGATGAGATCTGTGCGGCAAGAGGCGCATCGGCACTGTGATG 600

QY 607 GGGCCCCACCCCGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGACCCCTG 666
DB 601 GGGCCCCACCCCGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGACCCCTG 660
QY 667 AACTTCCCATGAGCCCATGAGACCGTGTGAGGCTGAGAGTGAAGCCCGGCAATGAGCGCG 726
DB 661 AACTTCCCATGAGCCCATGAGACCGTGTGAGGCTGAGAGTGAAGCCCGGCAATGAGCGCG 720
QY 727 CCCAAGGTGAGCAAGTGGCCCTGTGACCGAGGAAAGATCAAGGCGCTGACCGCATCTGCG 786
DB 721 CCCAAGGTGAGCAAGTGGCCCTGTGACCGAGGAAAGATCAAGGCGCTGACCGCATCTGCG 780
QY 787 GAGGAGTGAAGAGAGGGCGAAGATTCACCAAGATCGGCGCGGAGAACCCCTTCAACACCC 846
DB 781 GAGGAGTGAAGAGAGGGCGAAGATTCACCAAGATCGGCGCGGAGAACCCCTTCAACACCC 840
QY 847 CCGGTGTTCCCATGAGAGAGAGAGCAGACCAAGTGGCGCAAGCTGTGATCTTCCCGC 906
DB 841 CCGGTGTTCCCATGAGAGAGAGAGCAGACCAAGTGGCGCAAGCTGTGATCTTCCCGC 900
QY 907 GAGCTGAACAGCGCACCCAGCATCTTGTGGAGGTGCACTGGGCGATCCCCACCCCGCC 966
DB 901 GAGCTGAACAGCGCACCCAGCATCTTGTGGAGGTGCACTGGGCGATCCCCACCCCGCC 960
QY 967 GGGCTGAAGAGAGAGAGCGGTGACCGGTGTGAGCGTGGGCGAGCGCTTCAAGCGGTG 1026
DB 961 GGGCTGAAGAGAGAGAGCGGTGACCGGTGTGAGCGTGGGCGAGCGCTTCAAGCGGTG 1020
QY 1027 CCCCTGAGACGAGGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAG 1086
DB 1021 CCCCTGAGACGAGGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAG 1080
QY 1087 ACCCGCGCATCGCTTACCAAGTACAACTGTGCTGCGCCAGGGCTGAGAGGGCCAGCC 1146
DB 1081 ACCCGCGCATCGCTTACCAAGTACAACTGTGCTGCGCCAGGGCTGAGAGGGCCAGCC 1140
QY 1147 ATCTTCCAGAGAGAGATGACCAAGATCTCTGAGAGCCCTTCCGGCGCGCAACCCCGAGATC 1206
DB 1141 ATCTTCCAGAGAGAGATGACCAAGATCTCTGAGAGCCCTTCCGGCGCGCAACCCCGAGATC 1200
QY 1207 GTGATCTACAGATACATGAGACGACTGTAGTGGGCGAGCGACTGAGATTCGGCGAGCAC 1266
DB 1201 GTGATCTACAGATACATGAGACGACTGTAGTGGGCGAGCGACTGAGATTCGGCGAGCAC 1260
QY 1267 CGCGCCAGAGATGAGAGAGTGGCGCAAGCACCTGTGCGTGGGCGCTTACCAACCCCGCAGC 1326
DB 1261 CGCGCCAGAGATGAGAGAGTGGCGCAAGCACCTGTGCGTGGGCGCTTACCAACCCCGCAGC 1320
QY 1327 AAGAAACACCAAGAGAGCCCCCTTCTGTGTGATGGGCTTACGAGTGGACCCCGCAAG 1386
DB 1321 AAGAAACACCAAGAGAGCCCCCTTCTGTGTGATGGGCTTACGAGTGGACCCCGCAAG 1380
QY 1387 TGAACCGTGCAGCCCATGAGGCTGCGCGGAGAGAGAGCTGAGACGCTGAACGATCCAG 1446
DB 1381 TGAACCGTGCAGCCCATGAGGCTGCGCGGAGAGAGAGCTGAGACGCTGAACGATCCAG 1440
QY 1447 AAGCTGTGTGGCAAGTGAATGTGGCGAGCCAGATCTTACCCCGCATCAAGGTGGCGCAG 1506
DB 1441 AAGCTGTGTGGCAAGTGAATGTGGCGAGCCAGATCTTACCCCGCATCAAGGTGGCGCAG 1500
QY 1507 CTGTGCAAGCTGTGTGCGGCGCGCAAGGCGCTGACCGCAATGTGCGCTTGAACCGAGAG 1566
DB 1501 CTGTGCAAGCTGTGTGCGGCGCGCAAGGCGCTGACCGCAATGTGCGCTTGAACCGAGAG 1560
QY 1567 GCGGAGCTGAGAGTGGCGAGAACCGCGAGATCTTGTGCGGAGCGGTGCAACGCGGTGAC 1626
DB 1561 GCGGAGCTGAGAGTGGCGAGAACCGCGAGATCTTGTGCGGAGCGGTGCAACGCGGTGAC 1620
QY 1627 TACGACCCCAAGAGACCTGTGTGCGAGATTCAGAAACAGAGGCGACGACGATGTGACCC 1686
DB 1621 TACGACCCCAAGAGACCTGTGTGCGAGATTCAGAAACAGAGGCGACGACGATGTGACCC 1680
QY 1687 TACGAGATCTACAGAGAGCCCTTCAAGAACTGTGAAGACCGGCGCAAGTACCCCAAGATGCC 1746

Db 541 CAGTACAGCAGATCTGATCGAGATCTGCGCAAGAGCCATCGGACCGTGTGATC 600
Qy 607 GGGCCCCACCCCGTGAACATCATCGGCGCAACATCTGACCCAGCTGGAGTCAACCTG 666
Db 601 GGGCCCCACCCCGTGAACATCATCGGCGCAACATCTGACCCAGCTGGAGTCAACCTG 660
Qy 667 AACCTTCCCATCAGCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATGACGGC 726
Db 661 AACCTTCCCATCAGCCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATGACGGC 720
Qy 727 CCCAAGGTGAAGAGTGGCCCCCTGACCGAGAGAAATATCAAGGCCCTGACCCGCATCTGC 786
Db 721 CCCAAGGTGAAGAGTGGCCCCCTGACCGAGAGAAATATCAAGGCCCTGACCCGCATCTGC 780
Qy 787 GAGAGATGAGAGAGAGAGGAGCAAGATCAACAAGTGGCCCCGAGAAACCTTCAACACC 846
Db 781 GAGAGATGAGAGAGAGAGGAGCAAGATCAACAAGTGGCCCCGAGAAACCTTCAACACC 840
Qy 847 CCGGTGTTCGCATCAAGAAAGAGCAAGCAACAAGTGGCGCAAGCTGTGATCTTCCGC 906
Db 841 CCGGTGTTCGCATCAAGAAAGAGCAAGCAACAAGTGGCGCAAGCTGTGATCTTCCGC 900
Qy 907 GAGCTGAACAAGGCAACCCAGGACTTCTGGAGAGTCAAGTGGGCAATCCCCCAACCCGCC 966
Db 901 GAGCTGAACAAGGCAACCCAGGACTTCTGGAGAGTCAAGTGGGCAATCCCCCAACCCGCC 960
Qy 967 GGCCTGAAGAAAGAAAGAGCGTGAACCGTGCCTGAGAGTGGGCGAACGCTTCAAGCGT 1026
Db 961 GGCCTGAAGAAAGAAAGAGCGTGAACCGTGCCTGAGAGTGGGCGAACGCTTCAAGCGT 1020
Qy 1027 CCCCTGAGCAGAGACTTCCGCAAGTACACCGCCTTCAACATCCCAAGCATCAACAGAG 1086
Db 1021 CCCCTGAGCAGAGACTTCCGCAAGTACACCGCCTTCAACATCCCAAGCATCAACAGAG 1080
Qy 1087 ACCCCCGGCAATCCGCTACCAAGTCAACAGTGTCTGCCCCAGGAGCTGGAAGGGCAACCCCAAC 1146
Db 1081 ACCCCCGGCAATCCGCTACCAAGTCAACAGTGTCTGCCCCAGGAGCTGGAAGGGCAACCCCAAC 1140
Qy 1147 ATCTTCAAGAGCAGCAATGACCAAGATCTTGAAGCCCTTCCGCGCCGCAACCCCGAAGTC 1206
Db 1141 ATCTTCAAGAGCAGCAATGACCAAGATCTTGAAGCCCTTCCGCGCCGCAACCCCGAAGTC 1200
Qy 1207 GTGATTAACCAAGTACATGACGACTGTACGTGGAGCAGCACTTGAAGATCGGACGAC 1266
Db 1201 GTGATTAACCAAGTACATGACGACTGTACGTGGAGCAGCACTTGAAGATCGGACGAC 1260
Qy 1267 CCGCGCAAGATCGAGAGGCTGCGCAAGCACCTGTGCTGGGGCTTCAACACCCCGAC 1326
Db 1261 CCGCGCAAGATCGAGAGGCTGCGCAAGCACCTGTGCTGGGGCTTCAACACCCCGAC 1320
Qy 1327 AAGAGCAACCAAGAGAGGCCCTTCTGTGATGGGGCTACGAGGCTGCAACCCCGAAG 1386
Db 1321 AAGAGCAACCAAGAGAGGCCCTTCTGTGATGGGGCTACGAGGCTGCAACCCCGAAG 1380
Qy 1387 TGAACCGTGAAGCCCATCGAGCTGCCCCGAGAGAGAGTGAACCGTGAACGACATCCAG 1446
Db 1381 TGAACCGTGAAGCCCATCGAGCTGCCCCGAGAGAGAGTGAACCGTGAACGACATCCAG 1440
Qy 1447 AAGCTGTGGGCAAGCTGAACCTGGGCGCAAGCATTAACCCCGCATCAAGGTGGCGAG 1506
Db 1441 AAGCTGTGGGCAAGCTGAACCTGGGCGCAAGCATTAACCCCGCATCAAGGTGGCGAG 1500
Qy 1507 CTGTGCAAGCTGTGAGCGGCGCAAGGCCCTGACCGAGCATCGTGGCCCTGACCGAGAG 1566
Db 1501 CTGTGCAAGCTGTGAGCGGCGCAAGGCCCTGACCGAGCATCGTGGCCCTGACCGAGAG 1560
Qy 1567 GCCGAGCTGAGGCTGGCCGAGAACCGGAGATCTTGGCGAGGCCCGTGAACCGGCTGTAC 1626
Db 1561 GCCGAGCTGAGGCTGGCCGAGAACCGGAGATCTTGGCGAGGCCCGTGAACCGGCTGTAC 1620
Qy 1627 TACGACCCGAGCAAGGACTGTGTGGCGAGATCCAGAGCAAGGCCGACGACAGTGTGACC 1686

Db 1621 TACGACCCGAGCAAGGACTGTGTGGCGAGATCCAGAGCAGAGGCCGACGACCAAGTGAAC 1680
Qy 1687 TACGAGATTAACAGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCCAAGTCCG 1746
Db 1681 TACGAGATTAACAGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCCCAAGTCCG 1740
Qy 1747 ACCGCGCAACCAAGAGAGCTGAAGAGCTGAACCGAGGGCGGTCAAGAAATCCGCATGGAG 1806
Db 1741 ACCGCGCAACCAAGAGAGCTGAAGAGCTGAACCGAGGGCGGTCAAGAAATCCGCATGGAG 1800
Qy 1807 AGCATGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGACTTGA 1866
Db 1801 AGCATGTGATCTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGACTTGA 1860
Qy 1867 GAGACCTGTGAGACCGACTTACTGAGAGGCACTGTGATCCCGAGTGGAGTTCTGTAAAC 1926
Db 1861 GAGACCTGTGAGACCGACTTACTGAGAGGCACTGTGATCCCGAGTGGAGTTCTGTAAAC 1920
Qy 1927 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGTGAGAAAGAGGCCCATTCGCGCGCGAG 1986
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGTGAGAAAGAGGCCCATTCGCGCGCGAG 1980
Qy 1987 ACCCTTCAAGTGAACGCGCGCAACCGGCAAGCAAGATCGGCAAGGCCGCTAGT 2046
Db 1981 ACCCTTCAAGTGAACGCGCGCAACCGGCAAGCAAGATCGGCAAGGCCGCTAGT 2040
Qy 2047 ACCGACCGGCGCGCGAGAGATCTGTAGCCTGACCGAGACCAACAAGAAAGCCGAG 2106
Db 2041 ACCGACCGGCGCGCGAGAGATCTGTAGCCTGACCGAGACCAACAAGAAAGCCGAG 2100
Qy 2107 CTGCAAGGCAATCAAGCTGGCCCTTGCAGAGCAAGCGGCAAGAGGTGAACATCTGTACCGAC 2166
Db 2101 CTGCAAGGCAATCAAGCTGGCCCTTGCAGAGCAAGCGGCAAGAGGTGAACATCTGTACCGAC 2160
Qy 2167 AGCAGGTACCGCTGGGCAATCAATCAAGGCCAGGCCCAAGAGAGGAGCGAGCTGAGT 2226
Db 2161 AGCAGGTACCGCTGGGCAATCAATCAAGGCCAGGCCCAAGAGAGGAGCGAGCTGAGT 2220
Qy 2227 AACCAAGTATCGAGAGCTGATCAAGAAAGAAAGTGTACTGTAGCTGGTGCCTGCC 2286
Db 2221 AACCAAGTATCGAGAGCTGATCAAGAAAGAAAGTGTACTGTAGCTGGTGCCTGCC 2280
Qy 2287 CACAGAGGCAATGGCGCGCAACGAGCATTCGACAGCTGTGAGCAAGGGCATCCGCAAG 2346
Db 2281 CACAGAGGCAATGGCGCGCAACGAGCATTCGACAGCTGTGAGCAAGGGCATCCGCAAG 2340
Qy 2347 GTGCTGTCTGTGAAGGCAATCGATGGCGGCAATGTGATCAACAGTACATGACGACCTG 2406
Db 2341 GTGCTGTCTGTGAAGGCAATCGATGGCGGCAATGTGATCAACAGTACATGACGACCTG 2400
Qy 2407 TACGTGGGCAAGCGCGCCCTTAGAGATCAATTAAGCTTCCCGGGCTAGCAACCGGT 2463
Db 2401 TACGTGGGCAAGCGCGCCCTTAGAGATCAATTAAGCTTCCCGGGCTAGCAACCGGT 2457

RESULT 5
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
AC ABL39960;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR97SYM SEQ ID NO:31.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW Immunogenic; Immunisation; gag; pol; vif; vpr; tat; rev; env; nef;
KW Immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
XX
PN Synthetic.
PN WO200204493-A2.

QY 1621 GTGTACTAGACCCGACGAGGACCTGTGTGGCCGAGATTCAGAGACGAGGCGCACGACGAG 1680
 DB 1615 GTGTACTAGACCCGACGAGGACCTGTGTGGCCGAGATTCAGAGACGAGGCGCACGACGAG 1674
 QY 1681 TGGACCTACAGATGACGAGGAGGCGCTTCAAGAACCTGAGACGAGGACGAGGACGAGG 1740
 DB 1675 TGGACCTACAGATGACGAGGAGGCGCTTCAAGAACCTGAGACGAGGAGGCGCTTCAAGAACG 1734
 QY 1741 ATGCGACCGCCGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1800
 DB 1735 ATGCGACCGCCGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1794
 QY 1801 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCTTCCAGAGAGAG 1860
 DB 1795 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCTTCCAGAGAGAG 1854
 QY 1861 ACCGTGGAGAGCCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1920
 DB 1855 ACCGTGGAGAGCCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1914
 QY 1921 GTGAAACACCCGCGCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1980
 DB 1915 GTGAAACACCCGCGCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1974
 QY 1981 GCCGAGACCTTCTGACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2040
 DB 1975 GCCGAGACCTTCTGACGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2034
 QY 2041 TACGTGACCGAGCGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2100
 DB 2035 TACGTGACCGAGCGCGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2094
 QY 2101 ACCGAGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2160
 DB 2095 ACCGAGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2154
 QY 2161 ACCGAGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2220
 DB 2155 ACCGAGCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2214
 QY 2221 CTGTGTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2280
 DB 2215 CTGTGTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2274
 QY 2281 CCGGACCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2340
 DB 2275 CCGGACCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2334
 QY 2341 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2400
 DB 2335 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2394
 QY 2401 GACCTGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2460
 DB 2395 GACCTGTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2454
 QY 2461 GGTGAATTC 2469
 DB 2455 GGTGAATTC 2463

RESULT 6
 ADM73765
 ID ADM73765 standard; DNA; 2463 BP.
 AC ADM73765;
 XX 03-JUN-2004 (first entry)
 XX HIV-1 polynucleotide #8.
 DE HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
 XX HIV type C protein; immunostimulant.
 KM

XX Human immunodeficiency virus 1.
 OS US2003233961-A1.
 PN 04-DEC-2003.
 XX 05-JUL-2001; 2001US-00899575.
 PF 05-JUL-2000; 2000US-00610313.
 PR (MEGR/) MEGREDE J Z.
 PA (BARN/) BARNETT S W.
 PA (ENG/) ENGELBRECHT S.
 PA (RENS/) RENSBURG B J V.
 XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;
 DR WPI; 2004-060515/06.
 XX New expression cassette comprising a polynucleotide sequence encoding an
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
 PT immunization, generating of packaging cell lines or in producing HIV Type
 PT C proteins.
 XX Claim 1; SEQ ID NO 31; 160pp; English.
 PS The invention relates to an expression cassette comprising a
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
 CC also relates to a recombinant expression system for use in a host cell
 CC comprising an expression cassette, where the polynucleotide sequence
 CC further comprises control elements capable of driving expression in the
 CC selected host cell, a cell comprising an expression cassette where the
 CC polynucleotide sequence further comprises control elements compatible
 CC with the expression in the cell and a composition for generating an
 CC immunological response, comprising an expression cassette. The expression
 CC cassette and the methods of the invention are useful in eliciting an
 CC immune response, in DNA immunisation, in generation of packaging cell
 CC lines and in producing HIV Type C proteins. This sequence represents an
 CC HIV-1 polynucleotide of the invention.
 XX
 SO Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;
 Query Match 98.9%; Score 2442.2; DB 12; Length 2463;
 Best Local Similarity 99.6%; Pred. No. 1.7e-294;
 Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
 QY 1 GTGACGCGACCATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
 DB 1 GTGACGCGACCATGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 60
 QY 61 CGGAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 DB 61 CGGAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 QY 121 CACATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 DB 121 CACATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 QY 181 GGCACCAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
 DB 181 GGCACCAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
 QY 241 TTCCGCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 DB 241 TTCCGCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
 QY 301 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
 DB 301 AGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
 QY 361 GGCACCTGAACCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420


```

Db      1315  CCGGCAAGAGACACAGAGAGAGCCCTTCTCTGTGATGAGGCTACAGCTGACCC 1374
Qy      1381  GACAGTGGACCGTGAAGCCCATTCAGAGCTGACCGAGAGAGAGGCTGAGACG 1440
Db      1375  GACAGTGGACCGTGAAGCCCATTCAGAGCTGACCGAGAGAGAGGCTGAGACG 1434
Qy      1441  ATCCAGAACTGTGTGGCAAGCTGAATGAGGCAAGCAAGATCTACCCCGCATCA 1500
Db      1435  ATCCAGAACTGTGTGGCAAGCTGAATGAGGCAAGCAAGATCTACCCCGCATCA 1494
Qy      1501  CGCCAGCTGTGTGAAGCTGTGCGCGCGCGCAAGGCTTGAACCGAATCGTGC 1560
Db      1495  CGCCAGCTGTGTGAAGCTGTGCGCGCGCGCAAGGCTTGAACCGAATCGTGC 1554
Qy      1561  GAGAGGCGGAGCTGAGAGCTGAGCGAGAGACCGAGATCTGTGCGGAGCGGTG 1620
Db      1555  GAGAGGCGGAGCTGAGAGCTGAGCGAGAGACCGAGATCTGTGCGGAGCGGTG 1614
Qy      1621  GTGTACTAGACCCGAGCAAGAGCTGTGCGCGAGATCCAGAGCAAGGCGCAG 1680
Db      1615  GTGTACTAGACCCGAGCAAGAGCTGTGCGCGAGATCCAGAGCAAGGCGCAG 1674
Qy      1681  TGGACCTTACAGATCTACAGAGAGCCCTTCAAGAACTTGAAGACCGGCAAG 1740
Db      1675  TGGACCTTACAGATCTACAGAGAGCCCTTCAAGAACTTGAAGACCGGCAAG 1734
Qy      1741  ATGGCAGCGGCGCACCAAGAGAGAGAGAGAGCTGACCGAGGCGGTGAGAG 1800
Db      1735  ATGGCAGCGGCGCACCAAGAGAGAGAGAGAGCTGACCGAGGCGGTGAGAG 1794
Qy      1801  ATGAGAGAGATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCCATCCAG 1860
Db      1795  ATGAGAGAGATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCCATCCAG 1854
Qy      1861  ACCTGGAGAGACTGTGTGACCGACTTACTGTGGAGGCCACTTGATCCCGAG 1920
Db      1855  ACCTGGAGAGACTGTGTGACCGACTTACTGTGGAGGCCACTTGATCCCGAG 1914
Qy      1921  GTGAAACACCCCTGTGTGAGAGCTGTGTGACAGCTGTGAGAGAGAGCCAT 1980
Db      1915  GTGAAACACCCCTGTGTGAGAGCTGTGTGACAGCTGTGAGAGAGAGCCAT 1974
Qy      1981  GCCGAGACTTCTAGTGTGACCGGCGCGCAACCGGAGAGCAAGATCCGAGCG 2040
Db      1975  GCCGAGACTTCTAGTGTGACCGGCGCGCAACCGGAGAGCAAGATCCGAGCG 2034
Qy      2041  TACGTGACCGAGCGGCGCGCGCAAGAGATGTGAGCTTGAACCGAGCAAC 2100
Db      2035  TACGTGACCGAGCGGCGCGCGCGCAAGAGATGTGAGCTTGAACCGAGCAAC 2094
Qy      2101  ACCGAGCTGAGAGCATTCAGCTGAGCGCTGTGAGAGCAAGGAGAGTGAAC 2160
Db      2095  ACCGAGCTGAGAGCATTCAGCTGAGCGCTGTGAGAGCAAGGAGAGTGAAC 2154
Qy      2161  ACCGAGCTGAGAGCATTCAGCTGAGCGCTGTGAGAGCAAGGAGAGCGAG 2220
Db      2155  ACCGAGCTGAGAGCATTCAGCTGAGCGCTGTGAGAGCAAGGAGAGCGAG 2214
Qy      2221  CTGTGTGACAGATCATTCAGAGCTGATCAAGAGAGAGAGTGTACTTGAAG 2280
Db      2215  CTGTGTGACAGATCATTCAGAGCTGATCAAGAGAGAGAGTGTACTTGAAG 2274
Qy      2281  CCGGCGCAAGAGAGATTCGAGGAGAGAGAGAGATTCGAGCTGTGAGAG 2340
Db      2275  CCGGCGCAAGAGAGATTCGAGGAGAGAGAGAGAGATTCGAGCTGTGAGAG 2334
Qy      2341  CGCAAGTGTCTTCTGTGACCGGAGATTCAGTGGCGGATGTGATCTACAG 2400
Db      2335  CGCAAGTGTCTTCTGTGACCGGAGATTCAGTGGCGGATGTGATCTACAG 2394
Qy      2401  GACGTGTACGTGGGAGCGGCGCGCTTGAAGATGATTAAGCTTCCGCGGCT 2460

```

```

Db      2395  GACCTGTACGTGGGAGCGGCGCGCTTGAAGATCAATTAAGTTCCCGGGCTAG 2454
Qy      2461  GGT 2463
Db      2455  GGT 2457

RESULT 8
ADCI3265
ID      ADCI3265 standard; DNA; 2457 BP.
XX
AC      ADCI3265;
XX
DT      18-DEC-2003 (first entry)
XX
DE      DNA of HIV construct p2POL-opt-YM_C SEQ ID NO 44.
XX
KW      expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prot;
XX      Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS      Human immunodeficiency virus.
XX
PM      WO2003004620-A2.
XX
PD      16-JAN-2003.
XX
PF      05-JUL-2002; 2002MO-US021420.
XX
PR      05-JUL-2001; 2001US-0303192P.
XX      31-AUG-2001; 2001US-0316860P.
XX      16-JAN-2002; 2002US-0349871P.
XX
PA      (CHIR) CHIRON CORP.
XX      (UVEST-) UNIV STELLENBOSCH.
XX
PI      Zur Megede J, Barnette SW, Lian Y, Engelbrecht S, Van Rensburg BJ,
XX      WPI; 2003-221593/21.
XX
PT      New expression cassette comprising a polynucleotide sequence encoding a
XX      polypeptide including an HIV gag, Env, Int, Nef, p15RNaseH, Pol, Tat,
XX      Prot, or Rev polypeptide, useful for immunisation, or generating
XX      packaging cell lines.
XX
PS      Disclosure; Fig 41; 301pp; English.
XX
CC      The invention relates to a novel expression cassette comprising a
XX      polynucleotide sequence encoding a polypeptide including an HIV gag, Env,
XX      Int, Nef, p15RNaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX      expression cassette can be used to treat HIV type C by gene therapy or
XX      used in the development of a vaccine. The gene delivery vector is
XX      administered intramuscularly, intracutaneously, intranasally,
XX      subcutaneously, intradermally, transdermally, intravaginally,
XX      intrarectally, orally or intravenously. The expression cassette is useful
XX      for immunisation, generating packaging cell lines and producing HIV
XX      polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX      Type C related sequence of the invention.
XX
SQ      Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match      98.6%; Score 2434.6; DB 10; Length 2457;
Best Local Similarity 99.6%; Pred. NO.1.5e-293;
Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy      1      GTGAGCGCCACCATGGCGGAGGCGCATGAGCCAGGCGCAACGAGGCGCAATCTGATGCGAG 60
Db      1      GTGAGCGCCACCATGGCGGAGGCGCATGAGCCAGGCGCAACGAGGCGCAATCTGATGCGAG 60
Qy      61      CGCAGCACTTCAAGAGGCGCCCAAGGCGCATCATCAAGTGTCTTCACTGCGGCGCAAGAGGCGC 120
Db      61      CGCAGCACTTCAAGAGGCGCCCAAGGCGCATCATCAAGTGTCTTCACTGCGGCGCAAGAGGCGC 120
Qy      121      CACATGCGCCGCAACTGCGCGCGCCCGCAAGAGAGGCTGTGAGAGTGCAGCAAGAG 180

```


[illegible]

Db	1201	GAGATCGTGAATCTAACCA-----GGCCCCcctGTGAACGTGGGCAAGCCGACCTGGAGATGGCC	1254
QY	1261	CAGACACGCGCCCAAGATCTAGAGAGCTGCGCAAGCACTCTGCTGGCTGGGGCTTCCACCAACC	1320
Db	1255	CAGCACCGCGCGCAAGATCTAGAGAGCTGCGCAAGCACTCTGCTGGCTGGGGCTTCCACCAACC	1314
QY	1321	CCCGACAAAGAGCACAGAGAGAGCCCCCTTCTCTGTGGAATGGGCTTACAGCTGCAACCC	1380
Db	1315	CCCGACAAAGAGCACAGAGAGAGCCCCCTTCTCTGTGGAATGGGCTTACAGCTGCAACCC	1374
QY	1381	GACAAATGTGACCCGTGTGCAAGCCCATCTGAACTCTGCGGAGAGAGAGCTGGAACCTGTGAAGAC	1440
Db	1375	GACAAATGTGACCCGTGTGCAAGCCCATCTGAACTCTGCGGAGAGAGAGCTGGAACCTGTGAAGAC	1434
QY	1441	ATCCAGAAGCTGTGTGGGCAAGCTTGAACCTGGGCGCAGCCAGATCTTACCCCGGCATCAAGTGTG	1500
Db	1435	ATCCAGAAGCTGTGTGGGCAAGCTTGAACCTGGGCGCAGCCAGATCTTACCCCGGCATCAAGTGTG	1494
QY	1501	CGCCAGCTGTGCAAGCTGTCTGCGCGGGCGCAAGGCCCTTGACCGACATCTGTGCCCTTGACC	1560
Db	1495	CGCCAGCTGTGCAAGCTGTCTGCGCGGGCGCAAGGCCCTTGACCGACATCTGTGCCCTTGACC	1554
QY	1551	GAGAGAGCGCGAGCTGTGAGCTGGCGCGGAGAACCGCGAGATCTCTGGCGCGAGCCCGTGTCAAGGC	1620
Db	1555	GAGAGAGCGCGAGCTGTGAGCTGGCGCGGAGAACCGCGAGATCTCTGGCGCGAGCCCGTGTCAAGGC	1614
QY	1621	GTTGTACTTACGACCCCAAGCAAGACCTTGATGGCCAGATCTCAGAAAGCAAGGCGCAAGCAAG	1680
Db	1615	GTTGTACTTACGACCCCAAGCAAGACCTTGATGGCCAGATCTCAGAAAGCAAGGCGCAAGCAAG	1674
QY	1681	TGGAACCTTACCAAGATCTTACCAAGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAAG	1740
Db	1675	TGGAACCTTACCAAGATCTTACCAAGAGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAAG	1734
QY	1741	ATGGCGACACCGGCCCAACCAACAGCTGTAAGAGCTGTACCCAGAGGCGCTGTCAAGAAATCGCC	1800
Db	1735	ATGGCGACACCGGCCCAACCAACAGCTGTAAGAGCTGTACCCAGAGGCGCTGTCAAGAAATCGCC	1794
QY	1801	ATGGAGAGCAATCTGATCTGTGGGGCAAGAGCCCCCAAGTTCCGCTCTGCGCCATCTCAGAAAGAG	1866
Db	1795	ATGGAGAGCAATCTGATCTGTGGGGCAAGAGCCCCCAAGTTCCGCTCTGCGCCATCTCAGAAAGAG	1854
QY	1861	ACCTGTGGAGACCTGTGTGACCGGACTTACTGTGCAAGGCCACCTTGGATCCCGAGTGGGAGTTTC	1920
Db	1855	ACCTGTGGAGACCTGTGTGACCGGACTTACTGTGCAAGGCCACCTTGGATCCCGAGTGGGAGTTTC	1914
QY	1921	GTGAACACCCCCCTCTGTGTGAAGCTGTGTGTACAGCTGTGAGAGAGAGCCCATCATCTGGC	1986
Db	1915	GTGAACACCCCCCTCTGTGTGAAGCTGTGTGTACAGCTGTGAGAGAGAGCCCATCATCTGGC	1974
QY	1981	GCCGAGACCTTCTTACGTGTGAAGGGGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGGC	2044
Db	1975	GCCGAGACCTTCTTACGTGTGAAGGGGCGGCCCAACCGCGAGACCAAGATCGGCAAGGCGCGGC	2034
QY	2041	TACGTGACCGGACCGGGGCGCGCAGAGAGTCTGTAGCCTTGACCGGACCAACCAACCAAGAG	2100
Db	2035	TACGTGACCGGACCGGGGCGCGCAGAGAGTCTGTAGCCTTGACCGGACCAACCAACCAAGAG	2094
QY	2101	ACCGAGCTGTGACGGCCATCTCAAGCTGGCCCTGTGAGGACAGCGCGGCAAGAGGTGTACATCTGTG	2166
Db	2095	ACCGAGCTGTGACGGCCATCTCAAGCTGGCCCTGTGAGGACAGCGCGGCAAGAGGTGTACATCTGTG	2154
QY	2161	ACCGACAGCGATACGCGCTGGGGCATCATCCAGGCGCCAGCGCCGACCAAGAGGAGAGCGAG	2220
Db	2155	ACCGACAGCGATACGCGCTGGGGCATCATCCAGGCGCCAGCGCGCCGACCAAGAGGAGAGCGAG	2214
QY	2221	CTGTGTGAACCAAGATCATCTGAGCAGCTGTATCAAGAGAGAGAGAGGTGTACTTGAAGCTTGGGTG	2280
Db	2215	CTGTGTGAACCAAGATCATCTGAGCAGCTGTATCAAGAGAGAGAGAGGTGTACTTGAAGCTTGGGTG	2274
QY	2281	CCCGGCCCAAGAGGGGCACTGGGGCGGACAGAGCAATCCAGACGCTGTGTAGCAAGGGCACTC	2340
Db	2275	CCCGGCCCAAGAGGGGCACTGGGGCGGACAGAGCAATCCAGACGCTGTGTAGCAAGGGCACTC	2334

QY 2341 CGCAAGTCTCTTCTTCTGAGCGGCATGATGCGGATGATGATCTACAGATGAGC 2400
| | | | |
DB 2335 CGCAAGTCTCTTCTTCTGAGCGGCATGATGCGGATGATGATCTACAGATGAGC 2394
| | | | |
QY 2401 GACCTGTACGTGGGAGCGCGCGCTTACGATGATTAAGCTTCCCGGGCTAGCACC 2460
| | | | |
DB 2395 GACCTGTACGTGGGAGCGCGCGCTTACGATGATTAAGCTTCCCGGGCTAGCACC 2454
| | | | |
QY 2461 GGT 2463
| | | | |
DB 2455 GGT 2457
| | | | |
RESULT 9
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX ABL39961;
AC
XX
XX
DT 15-MAY-2002 (first entry)
XX
XX
DE Synthetic construct PR975YMM SEQ ID NO:32.
XX
XX
KM Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
XX immunostimulant; gene therapy; gene; de.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
XX
PR 05-JUL-2000; 2000US-00610313.
XX
XX
PA (CHTR) CHIRON CORP.
PA (UYST-) UNIV STELLERBOSCH.
XX
PI Zur Megele J, Barnett SM, Engelbrecht S, Van Rensburg EJ;
XX
XX
DR WPI; 2002-154920/20.
XX
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
XX applications including DNA immunisation or generation of packaging cell
XX lines, particularly in gene therapy.
XX
XX
PS Claim 1; Fig 10; 233pp; English.
XX
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (i) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention.
XX
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;
| | | | |
Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
Best Local Similarity 99.3%; Pred. No. 3.7e-291;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
| | | | |
QY 1 GTGACGCGCAATGCGCGAGCGCATGAGCCAGGCCAGGCCAGCCAACTCTGATGCG 60
| | | | |
DB 1 GTGACGCGCAATGCGCGAGCGCATGAGCCAGGCCAGGCCAGGCCAACTCTGATGCG 60
| | | | |

QY 61 CGCAGCACTTCAAGGGGCCCCAAGGCATCATCAAGTCTTCAACTGCGGCAAGAGGAGC 120
| | | | |
DB 61 CGCAGCACTTCAAGGGGCCCCAAGGCATCATCAAGTCTTCAACTGCGGCAAGAGGAGC 120
| | | | |
QY 121 CACATGCGCGCAACTGCGCGCGCCCCCGCAAGAGGCTGTGAGTGTGCGCAAGAG 180
| | | | |
DB 121 CACATGCGCGCAACTGCGCGCGCCCCCGCAAGAGGCTGTGAGTGTGCGCAAGAG 180
| | | | |
QY 181 GGCACCGAATGAAGATCTGACCGAGCGGCAAGCTTCTTCCGGAAGAGACTTGCC 240
| | | | |
DB 181 GGCACCGAATGAAGATCTGACCGAGCGGCAAGCTTCTTCCGGAAGAGACTTGCC 240
| | | | |
QY 241 TTCCCCCAGGCGAAGGCCCCGAGATTCCCCCAGCGAGCAAGAACCGGCCAACGCCACC 300
| | | | |
DB 241 TTCCCCCAGGCGAAGGCCCCGAGATTCCCCCAGCGAGCAAGAACCGGCCAACGCCACC 300
| | | | |
QY 301 AGCGCGAGCTGACAGTGTGCGCGGCAACCCCGGAGCGAGGCGCGCGCGAGCGCGAG 360
| | | | |
DB 301 AGCGCGAGCTGACAGTGTGCGCGGCAACCCCGGAGCGAGGCGCGCGCGAGCGCGAG 360
| | | | |
QY 361 GGCACCTTGAATCTTCCCCCAATCACTCTGTGAGCGGCCCCCTGTGTGATCAAGGTG 420
| | | | |
DB 361 GGCACCTTGAATCTTCCCCCAATCACTCTGTGAGCGGCCCCCTGTGTGATCAAGGTG 420
| | | | |
QY 421 GCGCGCAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGGTGTGAGAG 480
| | | | |
DB 421 GCGCGCAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGGTGTGAGAG 480
| | | | |
QY 481 ATGAGCTTGTCCCGCAATGGAAGCCCAAGATGATGTGCGCGCATGCGCGCTTCATCAG 540
| | | | |
DB 481 ATGAGCTTGTCCCGCAATGGAAGCCCAAGATGATGTGCGCGCATGCGCGCTTCATCAG 540
| | | | |
QY 541 GTGCGCCAGTACAGACCAATCTGTATCGAGATCTGCGGCAAGAGCCCATGCGACCGTG 600
| | | | |
DB 541 GTGCGCCAGTACAGACCAATCTGTATCGAGATCTGCGGCAAGAGCCCATGCGACCGTG 600
| | | | |
QY 601 CTGATCGGCGCCCAACCCCGGTGACATCATCGGCGGCAACATGCTGAGCCGTTGAGCGGC 660
| | | | |
DB 601 CTGATCGGCGCCCAACCCCGGTGACATCATCGGCGGCAACATGCTGAGCGGC 660
| | | | |
QY 661 ACCCTGAATCTTCCCGCATGAGCCCATGAGACCGGTGCGGTGAGCTGAGACCGCGCATG 720
| | | | |
DB 661 ACCCTGAATCTTCCCGCATGAGCCCATGAGACCGGTGCGGTGAGCTGAGACCGCGCATG 720
| | | | |
QY 721 GACGCGCCCAAGGTGAGAGATGAGGCGGCAAGATCAAGGCTTGAACCGGC 780
| | | | |
DB 721 GACGCGCCCAAGGTGAGAGATGAGGCGGCAAGATCAAGGCTTGAACCGGC 780
| | | | |
QY 781 ATCTGCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATGCGGCCCGAGAACCCCTTAC 840
| | | | |
DB 781 ATCTGCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATGCGGCCCGAGAACCCCTTAC 840
| | | | |
QY 841 AACACCCCGTGTGTCGATCAAGAGAGAGCAGACCAAGTGTGCGCAGCTGTGTGAGC 900
| | | | |
DB 841 AACACCCCGTGTGTCGATCAAGAGAGAGCAGACCAAGTGTGCGCAGCTGTGTGAGC 900
| | | | |
QY 901 TTCCGCGAGCTGAGCAAGCGCACCCAGAGACTTGTGAGGTGACCTGTGGCATCTCCCGAC 960
| | | | |
DB 901 TTCCGCGAGCTGAGCAAGCGCACCCAGAGACTTGTGAGGTGACCTGTGGCATCTCCCGAC 960
| | | | |
QY 961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGAACCGGTGAGAGGTGAGGAGCGCTTACTTTC 1020
| | | | |
DB 961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGAACCGGTGAGAGGTGAGGAGCGCTTACTTTC 1020
| | | | |
QY 1021 AGCGTGGCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAAC 1080
| | | | |
DB 1021 AGCGTGGCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAAC 1080
| | | | |
QY 1081 AACGAGACCCCGGCGATCGGCTTCAAGTACAACTGTGCGCCAGGCGTGAAGGCGCAGC 1140
| | | | |
DB 1081 AACGAGACCCCGGCGATCGGCTTCAAGTACAACTGTGCGCCAGGCGTGAAGGCGCAGC 1140
| | | | |

1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGGGCCGCAACCCC 1200
1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGGGCCGCAACCCC 1200
1201 GAGATCGTATCTACAGTACATGACAGCCTGTGAGTGGGAGGAGCACTGTGAGATCGGC 1260
1201 GAGATCGTATCTACCA-----GGCCCCCTGTGAGTGGGAGGAGCACTGTGAGATCGGC 1254
1261 CAGCACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGTGCGCTGGGGCTTCAACAC 1320
1255 CAGCACCGCGCCAGATCGAGGAGCTGCGCAAGCCTGTGCGCTGGGGCTTCAACAC 1314
1321 CCCGCAAGAGACCAAGAGAGCCCCCTTCTGTGATGGGCTACGAGCTTGACCCC 1380
1315 CCCGCAAGAGACCAAGAGAGCCCCCTTCTGTGATGGGCTACGAGCTTGACCCC 1368
1381 GACCAAGTGGACCGTGCAGCCCAATCGAGCTGCGCAAGAGAGAGCTGACCGTGAACGAC 1440
1369 GACCAAGTGGACCGTGCAGCCCAATCGAGCTGCGCAAGAGAGAGCTGACCGTGAACGAC 1428
1441 ATCCAGAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1500
1429 ATCCAGAGCTGTGGGCAAGCTGAATGGGCGCAGCAGATCTACCCCGGCATCAAGGTG 1488
1501 CCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCTTGAAC 1560
1489 CCGCAGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGACATCGTGCCTTGAAC 1548
1561 GAGGAGCGCGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGCGCGAGCGCGTGAAC 1620
1549 GAGGAGCGCGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGCGCGAGCGCGTGAAC 1608
1621 GTGTACTAGACCCCGCAAGAGCCTGTGCGCGAGATCTGAGAGAGGCGCAACGACGAC 1680
1609 GTGTACTAGACCCCGCAAGAGCCTGTGCGCGAGATCTGAGAGAGGCGCAACGACGAC 1668
1681 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCAAG 1740
1669 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGCGCAAGTACGCAAG 1728
1741 ATGGCCAGCGCCCAACCAAGCAGTGAAGAGCTGACCGAGGCGGTGAGAGAGATCGAC 1800
1729 ATGGCCAGCGCCCAACCAAGCAGTGAAGAGCTGACCGAGGCGGTGAGAGAGATCGAC 1788
1801 ATGAGAGCATCGTGTCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATTCAGAGAGAG 1860
1789 ATGAGAGCATCGTGTCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATTCAGAGAGAG 1848
1861 ACCCTGAGAGACCTGTGTGAACCGACTTCTGTGAGGAGCCACTGTATCCCGAGTGGAGTTTC 1920
1849 ACCCTGAGAGACCTGTGTGAACCGACTTCTGTGAGGAGCCACTGTATCCCGAGTGGAGTTTC 1908
1921 GTGAAACCCCCCTGTGTGAAGCTGTGTGACCACTGTGAGAGAGAGCCCAATCTGGC 1980
1909 GTGAAACCCCCCTGTGTGAAGCTGTGTGACCACTGTGAGAGAGAGCCCAATCTGGC 1968
1981 GCCGAGACCTTCTAGTGTGAACCGCGCCCAACCGGAGAGCAAGATCCGCAAGGCGCGC 2040
1969 GCCGAGACCTTCTAGTGTGAACCGCGCCCAACCGGAGAGCAAGATCCGCAAGGCGCGC 2028
2041 TACGTGACCGACCGGGGCGCGCAAGATCTGTGAGCCCTGACCGAGACCAACCAAGAG 2100
2029 TACGTGACCGACCGGGGCGCGCAAGATCTGTGAGCCCTGACCGAGACCAACCAAGAG 2088
2101 ACCGAGCTGAGGCGCATTCAGCTGTGCGCTGTGAGAGACGCGGCGAGGAGTGAACATCTGT 2160
2089 ACCGAGCTGAGGCGCATTCAGCTGTGCGCTGTGAGAGACGCGGCGAGGAGTGAACATCTGT 2148
2161 ACCGAGCGCGAGAGCGCTGGGCGCATTCAGGCGCGAGCGCGCAAGAGCGAGAGCGAG 2220
2149 ACCGAGCGCGAGAGCGCTGGGCGCATTCAGGCGCGAGCGCGCAAGAGCGAGAGCGAG 2208
2221 CTGGTGAACCAAGATCATTCAGAGAGCTGATCAAGAGAGAGGATGTACTTGAAGCTGGGTG 2280

2209 CTGGTGAACCAAGATCATTCAGAGAGCTGATCAAGAGAGAGGATGTACTTGAAGCTGGGTG 2268
2281 CCGCGCCCAAGAGGAGTGGGCGCAACGAGATCGAACAGCTGTGTGAGAGAGGCGCATC 2340
2269 CCGCGCCCAAGAGGAGTGGGCGCAACGAGATCGAACAGCTGTGTGAGAGAGGCGCATC 2328
2341 CGCAAGGTCGTGTTCTTGTGAACGCGCATTCATGGCGGCACTCGATCTTACCAATGAC 2400
2329 CGCAAGGTCGTGTTCTTGTGAACGCGCATTCATGGCGGCACTCGATCTTACCAATGAC 2388
2401 GACCTGTGAGTGGGAGGCGCGCCCTAGATCGATTAAAGCTTCCGCGGCTGACACC 2460
2389 GACCTGTGAGTGGGAGGCGCGCCCTAGATCGATTAAAGCTTCCGCGGCTGACACC 2448
2461 GGTGAATTC 2469
2449 GGTGAATTC 2457
Db 2449 GGTGAATTC 2457
RESULT 10
ADM73766
ID ADM73766 strand; DNA; 2457 BP.
XX
AC ADM73766;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #9.
XX
KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
OS Human immunodeficiency virus 1.
XX
PN US2003223961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGEDE J Z.
PA (BARNETT S W.
PA (ENGELBRECHT S.
PA (RENSBURG E J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX

Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 97.8%; Score 2415.4; DB 12; Length 2457;
 Best Local Similarity 99.3%; Pred. No. 3.7e-291;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 1 GTGACGCGCACCATGAGCGCGACATGAGCCAGGCGCACCGAGCGCACATCTCTGATGCG 60
 DB 1 GTGACGCGCACCATGAGCGCGACATGAGCCAGGCGCACCGAGCGCACATCTCTGATGCG 60
 QY 61 CGCAGCACTTTCAAGAGGCGCGCAAGCGCATATCAAGTGTCTTAATCTGCGGCAAGAGGCG 120
 DB 61 CGCAGCACTTTCAAGAGGCGCGCAAGCGCATATCAAGTGTCTTAATCTGCGGCAAGAGGCG 120
 QY 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 DB 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
 QY 181 GCGCACACAGATGAGAGCTGCAACCGAGCGCGCACAGCGCACTTCTTCCGCGAGGACCTTGGCC 240
 DB 181 GCGCACACAGATGAGAGCTGCAACCGAGCGCGCACAGCGCACTTCTTCCGCGAGGACCTTGGCC 240
 QY 241 TTCCCGCGAGGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
 DB 241 TTCCCGCGAGGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
 QY 301 AGCGCGAGCTGCAAGTGCAGCGCGCAACCGCGCGAGGAGGCGCGCGCGCGCGCGCGCGCG 360
 DB 301 AGCGCGAGCTGCAAGTGCAGCGCGCAACCGCGCGAGGAGGCGCGCGCGCGCGCGCGCGCG 360
 QY 361 GCGACCTTGAACTTCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 420
 DB 361 GCGACCTTGAACTTCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 420
 QY 421 GCGCGCGAGATTAAGAGGCGCGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 DB 421 GCGCGCGAGATTAAGAGGCGCGTGTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 QY 481 ATGAGCTGTGCGCGCGCAAGTGAAGCGCGCAAGATGATCGCGCGCGCGCGCGCGCGCGCG 540
 DB 481 ATGAGCTGTGCGCGCGCAAGTGAAGCGCGCAAGATGATCGCGCGCGCGCGCGCGCGCGCG 540
 QY 541 GTGCGCGAGTACAGACAGATCTGTATCGAGATCTGTGTGAGCGCGCGCGCGCGCGCGCG 600
 DB 541 GTGCGCGAGTACAGACAGATCTGTATCGAGATCTGTGTGAGCGCGCGCGCGCGCGCGCG 600
 QY 601 CTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 DB 601 CTGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 QY 661 ACCCTGAACCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
 DB 661 ACCCTGAACCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
 QY 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 DB 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
 QY 781 ATCTGCGAGAGATGAGAGAGGCGCGCAAGATCAACAGATGCGCGCGCGCGCGCGCGCG 840
 DB 781 ATCTGCGAGAGATGAGAGAGGCGCGCAAGATCAACAGATGCGCGCGCGCGCGCGCGCG 840
 QY 841 AACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
 DB 841 AACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
 QY 901 TTCCCGAGCTGAAACAAGCGCACCGAGCACTTCTGTGAGGTGCAAGTGTGCGCACTTCC 960
 DB 901 TTCCCGAGCTGAAACAAGCGCACCGAGCACTTCTGTGAGGTGCAAGTGTGCGCACTTCC 960
 QY 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 DB 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020

QY 1021 AGCGTGCCTTGTGACAGAGAGCTTCCGCAAGTACACCGCGCTTCAACCATCCCGAGCATCAAC 1080
 DB 1021 AGCGTGCCTTGTGACAGAGAGCTTCCGCAAGTACACCGCGCTTCAACCATCCCGAGCATCAAC 1080
 QY 1081 AACGAGACCCCGCGCATCCGCTTACAGATGACAGTGTGCTGCGCGCGCGCGCGCGCGCG 1140
 DB 1081 AACGAGACCCCGCGCATCCGCTTACAGATGACAGTGTGCTGCGCGCGCGCGCGCGCGCG 1140
 QY 1141 CCGAGCATCTTCCAGAGAGAGATGACAGAGATCTGTGAGCGCTTCCGCGCGCGCGCGCG 1200
 DB 1141 CCGAGCATCTTCCAGAGAGAGATGACAGAGATCTGTGAGCGCTTCCGCGCGCGCGCGCG 1200
 QY 1201 GAGATCGATCTTACAGTACATGACAGACCTGTGATGAGGCGCGCGCGCGCGCGCGCG 1260
 DB 1201 GAGATCGATCTTACAGTACATGACAGACCTGTGATGAGGCGCGCGCGCGCGCGCGCG 1260
 QY 1261 CAGCACCGCGCGCAAGATGAGAGAGCTGTGCGCAAGCACTGTGTGCGCTGTGCGCTTCA 1320
 DB 1261 CAGCACCGCGCGCAAGATGAGAGAGCTGTGCGCAAGCACTGTGTGCGCTGTGCGCTTCA 1320
 QY 1321 CCGGACAAAGAGACAGAGAGAGCGCGCGCTTCTGTGTGAGTGTGCTTACAGCTTCAAC 1380
 DB 1321 CCGGACAAAGAGACAGAGAGAGCGCGCGCTTCTGTGTGAGTGTGCTTACAGCTTCAAC 1380
 QY 1381 GACAAAGTGAACCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
 DB 1381 GACAAAGTGAACCGTGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
 QY 1441 ATCCAGAGCTGTGTGAGAGAGCTGAACTGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1500
 DB 1441 ATCCAGAGCTGTGTGAGAGAGCTGAACTGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1500
 QY 1501 GCGGAGCTGTGAGAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
 DB 1501 GCGGAGCTGTGAGAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
 QY 1561 GAGGAGCGCGAGCTGTGAGAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620
 DB 1561 GAGGAGCGCGAGCTGTGAGAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620
 QY 1621 GTGTACTTACGACCGCGCGCAAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
 DB 1621 GTGTACTTACGACCGCGCGCAAGAGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
 QY 1681 TGGACCTTACAGATTTTACAGAGAGCGCTTCAAGAACTGTGAGAGCGCGCGCAAGTAC 1740
 DB 1681 TGGACCTTACAGATTTTACAGAGAGCGCTTCAAGAACTGTGAGAGCGCGCGCAAGTAC 1740
 QY 1741 ATGCGCACCGCGCGCAACCAAGAGCTGTGAGAGAGCGCGCGCGCGCGCGCGCGCGCG 1800
 DB 1741 ATGCGCACCGCGCGCAACCAAGAGCTGTGAGAGAGCGCGCGCGCGCGCGCGCGCGCG 1800
 QY 1801 ATGAGAGCATGTGTATCTGTGTGCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860
 DB 1801 ATGAGAGCATGTGTATCTGTGTGCGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1860
 QY 1861 ACCTGTGAGAGCTGTGTGAGCGAGCTTACAGAGAGCGCGCGCGCGCGCGCGCGCGCG 1920
 DB 1861 ACCTGTGAGAGCTGTGTGAGCGAGCTTACAGAGAGCGCGCGCGCGCGCGCGCGCGCG 1920
 QY 1921 GTGAAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
 DB 1921 GTGAAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
 QY 1981 GCGGAGACCTTCTTACAGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
 DB 1981 GCGGAGACCTTCTTACAGTGAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2040
 QY 2041 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
 DB 2041 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
 QY 2029 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2088
 DB 2029 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2088

Db 841 CCCGTTCCCATCAAGAAAGAGACCAAGTGGCGAAGTGTGACCTTCCGC 900
QY 907 GAGCTGAACAAAGCCACCCAGAGACTTCTGAGAGTGCAGCTGGGCAATCCCCACCCCGC 966
Db 901 GAGCTGAACAAAGCCACCCAGAGACTTCTGAGAGTGCAGCTGGGCAATCCCCACCCCGC 960
QY 967 GGCCTGAAGAAGAAAGAGCGGTGACCGTGTGAGAGTGGGCGAAGCCTTCAAGCGGT 1026
Db 961 GGCCTGAAGAAGAAAGAGCGGTGACCGTGTGAGAGTGGGCGAAGCCTTCAAGCGGT 1020
QY 1027 CCCCTGAGAGAGACTTCTGAGAGTACACCGCTTTCACCATCCCAAGCAACAGAG 1086
Db 1021 CCCCTGAGAGAGACTTCTGAGAGTACACCGCTTTCACCATCCCAAGCAACAGAG 1080
QY 1087 ACCCCCGGATCCGCTTACCAAGTACAAAGTGTGCCCCAGAGGCTGAGAGGCGCCAGC 1146
Db 1081 ACCCCCGGATCCGCTTACCAAGTACAAAGTGTGCCCCAGAGGCTGAGAGGCGCCAGC 1140
QY 1147 ATCTTCCAGAGAGCAATGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCGAGATC 1206
Db 1141 ATCTTCCAGAGAGCAATGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCGAGATC 1200
QY 1207 GTGATCTACCAAGTACATGACCAAGTGTGAGAGCGCACTGAGATCCGCGAGAC 1266
Db 1201 GTGATCTACCA-----GGCCCCCTGTAGTGGGAGCGACCTGAGATCGGCGAGAC 1254
QY 1267 CGCGCCAGATGAGAGAGCTGCGCAAGCACTGTGCGCTGGGCGTTTCAACACCCCGAC 1326
Db 1255 CGCGCCAGATGAGAGAGCTGCGCAAGCACTGTGCGCTGGGCGTTTCAACACCCCGAC 1314
QY 1327 AAGAGCAACAGAGAGGCCCCCTTCTGTGAGTGGGCTAGAGCTGACCCCGAGCAAG 1386
Db 1315 AAGAGCAACAGAGAGGCCCCCTTCTGTGAGTGGGCTAGAGCTGACCCCGAGCAAG 1368
QY 1387 TGGACCTGTGAGAGCCCATGAGTGTGCGGAGAGAGAGCTGAGCCGTGACCAATCCAG 1446
Db 1369 TGGACCTGTGAGAGCCCATGAGTGTGCGGAGAGAGAGCTGAGCCGTGACCAATCCAG 1428
QY 1447 AAGCTGTGGGCAAGCTGAACTGGGCGAGCCAGATCTTACCCTGGCATCAAGTGGGCGAG 1506
Db 1429 AAGCTGTGGGCAAGCTGAACTGGGCGAGCCAGATCTTACCCTGGCATCAAGTGGGCGAG 1488
QY 1507 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCAAGCATGTGCGCTGACCGAGAG 1566
Db 1489 CTGTGCAAGCTGTGCGCGCGCGCGCAAGGCGCTGACCAAGCATGTGCGCTGACCGAGAG 1548
QY 1567 GCGGAGCTGAGAGCTGCGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACCGCGTGTAC 1626
Db 1549 GCGGAGCTGAGAGCTGCGCGAGAACCGCGAGATCTTGGCGAGCCCGTGCACCGCGTGTAC 1608
QY 1627 TACGACCCCGAGAGAGCTGTGTGGCGGAGATTCAGAGAGAGGCGCAAGCATGTGAGAC 1686
Db 1609 TACGACCCCGAGAGAGCTGTGTGGCGGAGATTCAGAGAGAGGCGCAAGCATGTGAGAC 1668
QY 1687 TACGAGATCTACAGAGAGCCCTTCAAGACTGAAGACCGGAGAGTACGCGCAAGTGGC 1746
Db 1669 TACGAGATCTACAGAGAGCCCTTCAAGACTGAAGACCGGAGAGTACGCGCAAGTGGC 1728
QY 1747 ACCGCGCAACCAAGAGAGCTGAGAGAGCTGACCGAGGCGGTGACAGAGATCCCATGAGAG 1806
Db 1729 ACCGCGCAACCAAGAGAGCTGAGAGAGCTGACCGAGGCGGTGACAGAGATCCCATGAGAG 1788
QY 1807 AGCATGTGTATCTGGGCGAGAGACCCCGAAGTTCCGCTGCGCATCCAGAGAGAGACTGG 1866
Db 1789 AGCATGTGTATCTGGGCGAGAGACCCCGAAGTTCCGCTGCGCATCCAGAGAGAGACTGG 1848
QY 1867 GAGACTGTGTGAGCGAGACTACTGAGAGAGGCACTGTGATCCCGAGTGGAGTTCGTGAGAC 1926
Db 1849 GAGACTGTGTGAGCGAGACTACTGAGAGAGGCACTGTGATCCCGAGTGGAGTTCGTGAGAC 1908
QY 1927 ACCCCCGCTGTGTGAGCTGTGTATCAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1986

Db 1909 ACCCCCCCTGTGTGAAGCTGTGTACCAAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1968
QY 1987 ACCTTCTACGTGAGATGGGCGCGCGCAACCGGAGACCAAGATGTGGCAAGGCGCGCTTACGTG 2046
Db 1969 ACCTTCTACGTGAGATGGGCGCGCGCAACCGGAGACCAAGATGTGGCAAGGCGCGCTTACGTG 2028
QY 2047 ACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGGAGACCAAGCAAGAGAGCGAG 2106
Db 2029 ACCGACCGGCGCGCGAGAGATCGTGAAGCTGACCGGAGACCAAGCAAGAGAGCGAG 2088
QY 2107 CTGCAAGCATTCAGCTGCGCTTGTGAGAGCAAGCGGCGAGAGTGAACAATGTGACCAAC 2166
Db 2089 CTGCAAGCATTCAGCTGCGCTTGTGAGAGCAAGCGGCGAGAGTGAACAATGTGACCAAC 2148
QY 2167 AGCGATAGCGCTGTGGGATCATTCAGGCGCGAGCCCGCAAGAGAGAGAGCGAGCTGTGG 2226
Db 2149 AGCGATAGCGCTGTGGGATCATTCAGGCGCGAGCCCGCAAGAGAGAGAGCGAGCTGTGG 2208
QY 2227 AACGATCATTCAGAGCACTGATCAAGAGAGAGAGTGTACTGAGCTGTGTGCGCGC 2286
Db 2209 AACGATCATTCAGAGCACTGATCAAGAGAGAGAGTGTACTGAGCTGTGTGCGCGC 2268
QY 2287 CACAAAGGCAATCGCGGCAACGAGAGATCGAACAAAGCTGTGAGCAAGGCAATCGCGAG 2346
Db 2269 CACAAAGGCAATCGCGGCAACGAGAGATCGAACAAAGCTGTGAGCAAGGCAATCGCGAG 2328
QY 2347 GTGCTGTCTGTGAAGCGCATGATGGCGGCAATCTGATCTTACAGTACATGAGCAACTGG 2406
Db 2339 GTGCTGTCTGTGAAGCGCATGATGGCGGCAATCTGATCTTACAGTACATGAGCAACTGG 2388
QY 2407 TACGTGGGCAAGCGGCGCGCTGAGATCGATTTAAAGCTTCCGCGGCTTACAGCCGT 2463
Db 2389 TACGTGGGCAAGCGGCGCGCTTACAGATCGATTTAAAGCTTCCGCGGCTTACAGCCGT 2445

RESULT 12
ADCI3264
ID ADCI3264 standard; DNA; 2445 BP.
XX
AC ADCI3264;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2POL-opt-YMM_C SEQ ID NO 43.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15Kaseh; Pol; Tat; Proct;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; da.
XX
OS Human immunodeficiency virus.
XX
PN MO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021420.
XX
PR 05-JUL-2001; 2001US-0301192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Llan Y, Engelbrecht S, Van Rensburg BJ,
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15Kaseh, Pol, Tat,
PT Proct, or Rev polypeptide, useful for immunisation, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 40; 301pp; English.

XX The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15maeh, Pol, Tat, Pro, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramusosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunization, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 97.3%; Score 2401.8; DB 10; Length 2445;
Best Local Similarity 99.2%; Pred. No. 1.8e-289;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```

OY 7 GCCACATGAGCGCGCATGAGCCAGGCGCACAGGCCCAACATCTGATGCAAGCGCAGC 66
DB 1 GCCACATGAGCGCGCATGAGCCAGGCGCACAGGCCCAACATCTGATGCAAGCGCAGC 60
OY 67 AACTTCAGAGGGCGCCAGAGGCAATCAATGCTTTCAATGCGGCGAAGAGGGCCACATC 126
DB 61 AACTTCAGAGGGCGCCAGAGGCAATCAATGCTTTCAATGCGGCGAAGAGGGCCACATC 120
OY 127 GCCCGCAATGCGCGCGCGCCCGCGCAAGAGGGCTGTGAGAGTGCGCGGCAAGAGGGCCAC 186
DB 121 GCCCGCAATGCGCGCGCGCCCGCGCAAGAGGGCTGTGAGAGTGCGCGGCAAGAGGGCCAC 180
OY 187 CAGATGAGAGATGCAACCGAGCGCCAGGCGCAATCTTCCGCGAGGAGACTGTGCTTCCCG 246
DB 181 CAGATGAGAGATGCAACCGAGCGCCAGGCGCAATCTTCCGCGAGGAGACTGTGCTTCCCG 240
OY 247 CAGGGCGAAGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAACAGCGCCCGC 306
DB 241 CAGGGCGAAGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAACAGCGCCCGC 300
OY 307 GAGCTGCAAGTGTGCGCGCGCAACCGCGCGAGGAGCGCGCGCGAGCGCGCGAGCGCGC 366
DB 301 GAGCTGCAAGTGTGCGCGCGCAACCGCGCGAGGAGCGCGCGCGCGAGCGCGCGAGCGCGC 360
OY 367 CTGAACTTCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCTGTGTGAGCATCAAGGTGTG 426
DB 361 CTGAACTTCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCTGTGTGAGCATCAAGGTGTG 420
OY 427 CAGATCAAGAGAGCGCTGTGTGAGCAACCGCGCGAGCAGCAACCTGTGTGAGAGATGAGC 486
DB 421 CAGATCAAGAGAGCGCTGTGTGAGCAACCGCGCGAGCAGCAACCTGTGTGAGAGATGAGC 480
OY 487 CTGCGCGGCAAGTGTGAGAGCGCGAGATGATGTGCGGAGTGTGCGGCGCTTCAATCAAGTGTG 546
DB 481 CTGCGCGGCAAGTGTGAGAGCGCGAGATGATGTGCGGAGTGTGCGGCGCTTCAATCAAGTGTG 540
OY 547 CAGTACGACAGAGATCTGTATGTGAGATCTGTGCGCAAGAGGCGCATCTGCGATCTGATC 606
DB 541 CAGTACGACAGAGATCTGTATGTGAGATCTGTGCGCAAGAGGCGCATCTGCGATCTGATC 600
OY 607 GGGCCCAACCGCGGTGAAATCATCTGCGCGCAATCTGCAATCTGCAATCTGCAATCTG 666
DB 601 GGGCCCAACCGCGGTGAAATCATCTGCGCGCAATCTGCAATCTGCAATCTGCAATCTG 660
OY 667 AACTTCCCATGTGCGCGCGCATCTGAGACCGGTGCGGTGAGTGAAGCGCGCGCATGTGAG 726
DB 661 AACTTCCCATGTGCGCGCGCATCTGAGACCGGTGCGGTGAGTGAAGCGCGCGCATGTGAG 720
OY 727 CCCAAGGTGAGAGATGTGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGTGACCGCATCTG 786
DB 721 CCCAAGGTGAGAGATGTGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGTGACCGCATCTG 780
OY 787 GAGAGATGTGAGAGAGAGGCGCAAGATCAACAGATGTGCGCGCGAGAACCTCTTCAACATCC 846
DB 787 GAGAGATGTGAGAGAGAGGCGCAAGATCAACAGATGTGCGCGCGAGAACCTCTTCAACATCC 846
```

```

DB 781 GAGAGATGTGAGAGAGAGGCGCAAGATCAACAGATGTGCGCGCGAGAACCTCTTCAACATCC 840
OY 847 CCCGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
DB 841 CCCGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
OY 907 GAGCTGAAACAAGCGGACCGCAGAGATCTTGTGGAGGTGAGCTGTGGAGATCCCGACCGCGC 966
DB 901 GAGCTGAAACAAGCGGACCGCAGAGATCTTGTGGAGGTGAGCTGTGGAGATCCCGACCGCGC 960
OY 967 GGCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
DB 961 GGCCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
OY 1027 CCCCTGAGAGAGAGATCTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
DB 1021 CCCCTGAGAGAGAGATCTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
OY 1087 ACCCGCGGAGTCCGCTTACAGTACAAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
DB 1081 ACCCGCGGAGTCCGCTTACAGTACAAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
OY 1147 ATCTTCCAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCCCGAGATC 1206
DB 1141 ATCTTCCAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCAACCCCGAGATC 1200
OY 1207 GTGATCTACAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
DB 1201 GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
OY 1267 CGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
DB 1255 CGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1314
OY 1327 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
DB 1315 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368
OY 1387 TGAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
DB 1369 TGAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428
OY 1447 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
DB 1429 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1488
OY 1507 CTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
DB 1489 CTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1548
OY 1567 GCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
DB 1549 GCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
OY 1627 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
DB 1609 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1668
OY 1687 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
DB 1669 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
OY 1747 ACCGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806
DB 1729 ACCGCGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1788
OY 1807 AGCATGTGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
DB 1789 AGCATGTGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848
OY 1867 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
DB 1849 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1908
```



```
OY 1927 ACCCCCCCTGTGAAGCTGTGTACACAGCTGAGAAAGAGCCCATCATCGGCGCGAG 1986
XX |||||
DB 1909 ACCCCCCCTGTGTGAAGCTGTGTACACAGCTGAGAAAGAGCCCATCATCGGCGCGAG 1968
OY 1987 ACCCTTCTAGTGAAGCGCGCGCCCAACCGCGAGACCAAGATCGGAGCGCGCTTACGTTG 2046
XX |||||
DB 1969 ACCCTTCTAGTGAAGCGCGCGCCCAACCGCGAGACCAAGATCGGAGCGCGCTTACGTTG 2028
OY 2047 ACCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2106
XX |||||
DB 2029 ACCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2088
OY 2107 CTGCAAGCGCATCAGAGTGGCCCTGTGACGACGACGCGGAGAGTGAACATCGTGAACCGAG 2166
XX |||||
DB 2089 CTGCAAGCGCATCAGAGTGGCCCTGTGACGACGACGCGGAGAGTGAACATCGTGAACCGAG 2148
OY 2167 AGCCAGTACGCGCTGTGGGATCATTCAGAGCCGACGCGCGACAGAGCGAGCGTGTG 2226
XX |||||
DB 2149 AGCCAGTACGCGCTGTGGGATCATTCAGAGCCGACGCGCGACAGAGCGAGCGTGTG 2208
OY 2227 AACCGATATCGACAGCTGTATCAAGAAAGAGAGTGTACTTGAAGTGGTGGCGCGC 2286
XX |||||
DB 2209 AACCGATATCGACAGCTGTATCAAGAAAGAGAGTGTACTTGAAGTGGTGGCGCGC 2268
OY 2287 CACAAGGCGCATCGCGCGCGCAACGACGATCGAACAGCTGTGAGCAAGGCGCATCGCGCAAG 2346
XX |||||
DB 2269 CACAAGGCGCATCGCGCGCGCAACGACGATCGAACAGCTGTGAGCAAGGCGCATCGCGCAAG 2328
OY 2347 GTGCTGTCTGTGAAGCGCATCGATGGCGCGCATCGTGTATCTACAGTACATGAGCAAGCTG 2406
XX |||||
DB 2329 GTGCTGTCTGTGAAGCGCATCGATGGCGCGCATCGTGTATCTACAGTACATGAGCAAGCTG 2388
OY 2407 TACGTTGGGAGAGGCGCGCGCGCTTACGATTCGATTTAAAGCTTCCCGGGGCTTACGACCGGT 2463
XX |||||
DB 2389 TACGTTGGGAGAGGCGCGCGCGCTTACGATTCGATTTAAAGCTTCCCGGGGCTTACGACCGGT 2445
```

RESULT 13
ADCl3230
ID ADCl3230 standard, DNA, 3930 BP.

```
XX ADCl3230;  
AC 18-DEC-2003 (first entry)  
DT  
XX  
DE DNA of HIV construct GagComp1Polmut_C SEQ ID NO 9.  
XX  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prot;  
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.  
XX  
OS Human immunodeficiency virus.  
XX  
PN WO2003004620-A2.  
XX  
PD 16-JAN-2003.  
XX  
PE 05-JUL-2002; 2002WO-US021420.  
XX  
XX 05-JUL-2001; 2001US-0303192P.  
PR 31-AUG-2001; 2001US-0316860P.  
PR 16-JAN-2002; 2002US-0349871P.  
XX  
PA (CHIR ) CHIRON CORP.  
XX (U1ST-) UNIV STELLENBOSCH.  
PI Zur Megele J, Barnett SW, Llan Y, Engelbrecht S, Van Renburg RJ,  
XX MPI, 2003-221593/21.  
DR  
XX  
PT New expression cassette comprising a polynucleotide sequence encoding a  
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RNaseH, Pol, Tat,  
PT Prot, or Rev polypeptide, useful for immunization, or generating
```

PT packaging cell lines.

XX
XX Disclosure; Fig 6; 301pp; English.

CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RNaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intradermally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX
SQ Sequence 3930 BP, 890 A, 1365 C, 1214 G, 461 T, 0 U, 0 Other;

Query Match 97.0%; Score 2394.8; DB 10; Length 3930;

Best Local Similarity 99.2%; Pred. No. 1.2e-288;

Matches 2431; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
OY 14 TGGCGAGGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 73
DB 1487 TGGCGAGGCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1546
OY 74 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 133
DB 1547 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1606
OY 134 ACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193
DB 1607 ACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1666
OY 194 AGGACTGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 253
DB 1667 AGGACTGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1726
OY 254 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 313
DB 1727 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1786
OY 314 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
DB 1787 AGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846
OY 374 TCCCGCGAGTACCGCTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
DB 1847 TCCCGCGAGTACCGCTGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906
OY 434 AGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
DB 1907 AGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
OY 494 GCAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 553
DB 1967 GCAAGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2026
OY 554 ACCAGATCTGTATGAGATCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
DB 2027 ACCAGATCTGTATGAGATCTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2086
OY 614 CCCCCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
DB 2087 CCCCCGTGAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2146
OY 674 CCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 733
DB 2147 CCATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2206
OY 734 TGAAGCAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 793
DB 2207 TGAAGCAATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2266
```

QY 794 TGGAGAGGAGGCAAGATCAACAAGATGGCCCCCGAGAACCCCTTACAACACCCCGGTGT 853
 Db 2267 TGGAGAGGAGGCAAGATCAACAAGATGGCCCCCGAGAACCCCTTACAACACCCCGGTGT 2326
 QY 854 TGGCCATCAAGAGAGAGAGCAGCAGCAAGTGGGCGCAAGTGTGACTTCCGCGAGCTGA 913
 Db 2327 TGGCCATCAAGAGAGAGAGCAGCAGCAAGTGGGCGCAAGTGTGACTTCCGCGAGCTGA 2386
 QY 914 ACAAGCGCACCCGAGACTTCTGGAGAGTGCAGCTGGGCAATCCCCCAACCCCGCGGCTGA 973
 Db 2387 ACAAGCGCACCCGAGACTTCTGGAGAGTGCAGCTGGGCAATCCCCCAACCCCGCGGCTGA 2446
 QY 974 AGAAGAGAGAGAGAGCTGACCTGTCTGAGAGTGGGCGACGCTTACTTCAAGCTGTCCCTGG 1033
 Db 2447 AGAAGAGAGAGAGAGCTGACCTGTCTGAGAGTGGGCGACGCTTACTTCAAGCTGTCCCTGG 2506
 QY 1034 ACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCCCGG 1093
 Db 2507 ACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAGAGACCCCGG 2566
 QY 1094 GCATCCGCTACCAAGTACAACTGTCTGAGAGTGGGCGTGGAGAGGCGAGCCCGAGCATCTTCC 1153
 Db 2567 GCATCCGCTACCAAGTACAACTGTCTGAGAGTGGGCGTGGAGAGGCGAGCCCGAGCATCTTCC 2626
 QY 1154 AGAGCAGCATGACCAAGATCTTGGAGAGCTTCCGCGCGCGCAACCCCGAGATGTGATCT 1213
 Db 2627 AGAGCAGCATGACCAAGATCTTGGAGAGCTTCCGCGCGCGCAACCCCGAGATGTGATCT 2686
 QY 1214 ACCAGTACATGAGCAGCTGTACGTGGGCGAGAGCTGGAGATGGGCGAGCACCGGCGCA 1273
 Db 2687 ACCA-----GAGCCCCCTGTACGTGGGCGAGAGCTGGAGATGGGCGAGCACCGGCGCA 2740
 QY 1274 AGATGAGAGAGCTGCGCAGAGCAGCTGTCTGTGGGCTTCAACACCCCGAGCAAGAG 1333
 Db 2741 AGATGAGAGAGCTGCGCAGAGCAGCTGTCTGTGGGCTTCAACACCCCGAGCAAGAG 2800
 QY 1334 ACCAGAGAGAGCCCCCTTCTGTGTGATGGGCTTACAGCTGACACCCCGAGAGTGGACCG 1393
 Db 2801 ACCAGAGAGAGCCCCCTTCTGTGTGATGGGCTTCTGTGTGATGGGCTTCTGTGTGATGGGCTT 2854
 QY 1394 TGGAGCCCATCGAGCTGCGCGAGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGCTGG 1453
 Db 2855 TGGAGCCCATCGAGCTGCGCGAGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGCTGG 2914
 QY 1454 TGGGCAAGCTGAATCTGGGCGAGCAGATCTTACCCCGGCAATCAAGTGTGCGCAGCTGTGCA 1513
 Db 2915 TGGGCAAGCTGAATCTGGGCGAGCAGATCTTACCCCGGCAATCAAGTGTGCGCAGCTGTGCA 2974
 QY 1514 AGCTGCTGCGCGCGCGCAGAGGCTCTTACCGACATCTGTGCTGACCGGAGAGAGGCGGAGC 1573
 Db 2975 AGCTGCTGCGCGCGCGCAGAGGCTCTTACCGACATCTGTGCTGACCGGAGAGAGGCGGAGC 3034
 QY 1574 TGGAGCTGGCGCGAGACCGGAGAGATCTTGGGCGAGACCGGCTGTGACCGGCTGTACTTACGACC 1633
 Db 3035 TGGAGCTGGCGCGAGACCGGAGAGATCTTGGGCGAGACCGGCTGTGACCGGCTGTACTTACGACC 3094
 QY 1634 CCAGCAAGAGCTGTGTGCGCGAGATCCAGAGAGAGGCGCAGACCAAGTGAACCTTACAGCA 1693
 Db 3095 CCAGCAAGAGCTGTGTGCGCGAGATCCAGAGAGAGGCGCAGACCAAGTGAACCTTACAGCA 3154
 QY 1694 TCTTACCAAGAGCTTCTTCAAGACCTTGAAGACCGGCGCAAGTACGCGCAAGATGCGACCGCC 1753
 Db 3155 TCTTACCAAGAGCTTCTTCAAGACCTTGAAGACCGGCGCAAGTACGCGCAAGATGCGACCGCC 3214
 QY 1754 ACACCAAGAGCTGTGTGAGCAGCTGACCGAGGCGGTGCGAGAGATGCGCATGAGAGAGATGCG 1813
 Db 3215 ACACCAAGAGCTGTGTGAGCAGCTGACCGAGGCGGTGCGAGAGATGCGCATGAGAGAGATGCG 3274
 QY 1814 TGAATCTGGGCGAGAGCCCGCAAGTTCGCTGTCCATCCAGAGAGAGACTTGGAGAGACT 1873
 Db 3275 TGAATCTGGGCGAGAGCCCGCAAGTTCGCTGTCCATCCAGAGAGAGACTTGGAGAGACT 3334

QY 1874 GGTGAGCCGACTTACCTGGCAGAGCCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGC 1933
 Db 3335 GGTGAGCCGACTTACCTGGCAGAGCCACCTGGATCCCGGAGTGGAGTTCGTGAACACCCCGC 3394
 QY 1934 CCTGTGAGAGCTGTGTGTACCAAGCTGTGAGAGAGAGCCATCATCGGCGCGAGACTTCT 1993
 Db 3395 CCTGTGAGAGCTGTGTGTACCAAGCTGTGAGAGAGAGCCATCATCGGCGCGAGACTTCT 3454
 QY 1994 AGTGTGAGCGGCGCGCGCAACCGCGAGACCAAGTGTGGCAAGGCGGCTTACGTGACCGACC 2053
 Db 3455 AGTGTGAGCGGCGCGCGCAACCGCGAGACCAAGTGTGGCAAGGCGGCTTACGTGACCGACC 3514
 QY 2054 GGGGCGCGAGAGATCGTGAAGCTGACCGGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAG 2113
 Db 3515 GGGGCGCGAGAGATCGTGAAGCTGACCGGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 3574
 QY 2114 CCATCCAGCTGGGCGGCTGTGAGAGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2173
 Db 3575 CCATCCAGCTGGGCGGCTGTGAGAGCAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3634
 QY 2174 AGGCGCTGGGCAATCATCCAGGCGCCAGACCCGAGACAGAGAGAGAGAGAGAGAGAGAGAG 2233
 Db 3635 AGGCGCTGGGCAATCATCCAGGCGCCAGACCCGAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 3694
 QY 2234 TCATGAGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2293
 Db 3695 TCATGAGAGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3754
 QY 2294 GCATCGGCGGCGCAAGCAGAGATGACCAAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
 Db 3755 GCATCGGCGGCGCAAGCAGAGATGACCAAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 3814
 QY 2354 TCTTGAAGCGGATCGATGAGGCGGATCTGTATCTTACCAATGAGAGAGAGAGAGAGAGAGAGAG 2413
 Db 3815 TCTTGAAGCGGATCGATGAGGCGGATCTGTATCTTACCAATGAGAGAGAGAGAGAGAGAGAGAG 3874
 QY 2414 GCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCGTGAACCGGCT 2463
 Db 3875 GCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCGTGAACCGGCT 3924

RESULT 14
 ADCL3231
 ID ADCL3231 standard; DNA; 3930 BP.
 XX
 AC ADCL3231;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE DNA of HIV construct GagComp1PolmutatC SEQ ID NO 10.
 XX
 KM expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prox;
 KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
 OS Human immunodeficiency virus.
 PN WO2003004620-A2.
 PD 16-JAN-2003.
 PF 05-JUL-2002; 2002WO-US021420.
 PR 05-JUL-2001; 2001US-0303192P.
 PR 31-AUG-2001; 2001US-031660P.
 PR 16-JAN-2002; 2002US-0349871P.
 PA (CHIR) CHIRON CORP.
 PA (OYST-) UNIV STIELEBROSCHE.
 PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ,
 PI PI
 DR WPI; 2003-221593/21.
 XX

Db 2147 CCATGAGCCCTTCAGAGCCGTGACCTGTAAGCCGGCATGAGCCGCCAAG 2206
Qy 734 TGAAGAGGCGCCCTGACCCGAGAGAGATGAAGCCCTTGAACCCCATCTGCGAGAGA 793
Db 2207 TGAAGAGGCGCCCTGACCCGAGAGAGATGAAGCCCTTGAACCCCATCTGCGAGAGA 2266
Qy 794 TGAAGAGGCGCGAAGATCACCAAGATCGGCCGAGAACCCCTTACAACACCCCGGT 853
Db 2267 TGAAGAGGCGCGAAGATCACCAAGATCGGCCGAGAACCCCTTACAACACCCCGGT 2326
Qy 854 TCGGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
Db 2327 TCGGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2386
Qy 914 ACAAGCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
Db 2387 ACAAGCGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2446
Qy 974 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
Db 2447 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
Qy 1034 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1093
Db 2507 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2566
Qy 1094 GCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
Db 2567 GCATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2626
Qy 1154 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1213
Db 2627 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2686
Qy 1214 ACCAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1273
Db 2687 ACCA-----GGCCCCCTGTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740
Qy 1274 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1333
Db 2741 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2800
Qy 1334 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1393
Db 2801 ACCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2854
Qy 1394 TGCAGCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1453
Db 2855 TGCAGCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2914
Qy 1454 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1513
Db 2915 TGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2974
Qy 1514 AGCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1573
Db 2975 AGCTGTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3034
Qy 1574 TGGAGCTGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1633
Db 3035 TGGAGCTGGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3094
Qy 1634 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1693
Db 3095 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3154
Qy 1694 TCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1753
Db 3155 TCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3214
Qy 1754 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1813

Db 3215 AACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3274
Qy 1814 TGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1873
Db 3275 TGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3334
Qy 1874 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1933
Db 3335 GGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3394
Qy 1934 CCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1993
Db 3395 CCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3454
Qy 1994 AGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2053
Db 3455 AGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3514
Qy 2054 GGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2113
Db 3515 GGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3574
Qy 2114 CCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2173
Db 3575 CCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3634
Qy 2174 AGGCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233
Db 3635 AGGCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3694
Qy 2234 TCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2293
Db 3695 TCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3754
Qy 2294 GCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
Db 3755 GCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3814
Qy 2354 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2413
Db 3815 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3874
Qy 2414 GCAGGCGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463
Db 3875 GCAGGCGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3924

Search completed: December 30, 2005, 08:56:53
Job time : 1308.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8718.85 Seconds
(without alignments) 13249.138 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtgcagccaccatgcccga.....gggctagcaccgggtgaattc 2469

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_esc1.*
2: gb_esc2.*
3: gb_esc3.*
4: gb_hic.*
5: gb_esc4.*
6: gb_esc5.*
7: gb_esc6.*
8: gb_esc7.*
9: gb_gsa1.*
10: gb_gsa2.*
11: gb_gsa3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	136.4	5.5	330	10	CL293849 02S0349-0
2	97.8	4.0	2031	10	CL974899 OaIFCC042
3	97.2	3.9	2886	10	CL967755 OaIFCC015
4	91.4	3.7	1509	10	CL959255 OaIFCC002
5	90.8	3.7	2598	4	AY103647 Zea mays
6	89.4	3.6	743	10	CZ247380 AIAA-aaf3
7	86.6	3.5	1398	10	CL961989 OaIFCC006
8	85.6	3.5	951	3	BM321451 rockefeller
c 9	85	3.4	869	7	CK159167 FGAS04056
10	84	3.4	3069	10	CL973991 OaIFCC025
11	83.8	3.4	1941	10	CL971508 OaIFCC021
12	82.6	3.3	1781	8	CK099623 RECM0260
c 13	80.8	3.3	1060	10	CK099623 RECM0260
14	80.6	3.3	892	10	CW922203 BDCAR29TR
15	80.2	3.2	1132	3	CZ216254 AIAA-aaf2
16	79.6	3.2	1165	3	BM320864 rockefeller
17	79	3.2	867	3	BM320900 rockefeller
18	78.4	3.2	11691	10	BM321430 rockefeller
19	78.2	3.2	1962	10	CL962901 OaIFCC008
20	78	3.2	2853	10	CL961326 OaIFCC005
21	78	3.2	1485	10	CL974397 OaIFCC025
22	76.6	3.1	1550	3	CL970981 OaIFCC020
					BM321022 rockefeller

23	76.6	3.1	2559	10	CL982027 OaIFSC046
24	76.2	3.1	1401	10	CL962721 OaIFSC038
25	75.6	3.1	671	6	CA093222 SCCCL300
26	75.4	3.1	2299	4	AY106831 Zea mays
27	75.2	3.0	854	7	CK777127 964131 MA
28	75	3.0	862	3	BM321023 rockefeller
29	75	3.0	3822	10	CL972913 OaIFSC023
30	74.6	3.0	914	9	BZ568300 PAC82-164
31	74.6	3.0	2010	10	CL962831 OaIFSC008
32	74.6	3.0	2313	10	CL982362 OaIFSC047
33	74.6	3.0	2682	10	CL969033 OaIFSC017
34	74.4	3.0	3134	4	AY109500 Zea mays
35	74.2	3.0	1680	10	CL982770 OaIFSC049
36	74	3.0	1290	10	CL972679 OaIFSC023
37	74	3.0	1386	11	DQ045165 Homo sapi
38	74	3.0	2072	4	CR603312 full-1eng
c 39	73.8	3.0	889	7	CK159613 OaIFSC047
c 40	73.6	3.0	757	9	CC678788 OGMI171TV
41	73.6	3.0	2028	10	CL979437 OaIFSC033
42	73.2	3.0	853	3	BM321393 rockefeller
43	73	3.0	1689	10	CL972373 OaIFSC022
c 44	72.8	2.9	566	3	BM587428 OaIFSC045
45	72.8	2.9	1575	10	CL979927 OaIFSC045

ALIGNMENTS

RESULT 1
LOCUS CL293849 330 bp DNA linear GSS 12-FRB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MutTail Library Zea mays genomic clone
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 330)
REFERENCE Lathaw,S., Tan,B.C., Settles,A.M. and McCarty,D.R.
AUTHORS Sequence tagged transposon insertions from the UniformMu maize population
TITLE Unpublished (2003)
JOURNAL Contact: Donald R. McCarty
COMMENT Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmcufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
02S0349-08, Primer, seq: A
Class: transposon insertion site.
Location/Qualifiers
1..330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/db_xref="taxon:4577"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/clone_1ib="UniformMu MutTail Library"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sephadex 400 spin columns and cloned into the TOPO PCR4 vector."

FEATURES

source

ORIGIN

Query Match	5 %;	Score 136.4;	DB 10;	Length 330;
Best Local Similarity	66.4%;	Pred. No. 2.8e-14;		
Matches 211;	Conservative	0;	Mismatches 106;	Indels 1;
				Gaps 1
Qy	421	GGCGGCGAGATCAAGAGAGGCCCTCTGTCGACACCGCGCGCAGCAGACCGTGTGTGAGGAG	480	
Db	329	GGGGGGGCGAGCTGGAGGAAGCTTATTAGATACAGAGCGAGATGATACAGTATTAAAGAA	270	
Qy	481	ATGAGCCTGCGCCCGGCAAGTGGAGGCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG	540	
Db	269	ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGCAATTGGAGGTTTATCAAA	210	
Qy	541	GTGGGCGCAGTACGACCAATCCTGATTCGAGATCTCGCGCAAAAGGCCATTCGCGACCGTG	600	
Db	209	GTAAACAGTATGATTCAGGTACCCATAGAAATCTGTGGGCATTAAGCTATAGGTACGGTA	150	
Qy	601	CTGATTCGGAGCCCAACCCCGTGAAATCATTCGGCGCCCAACATGCTGACCCAGCTGGAGCTGC	660	
Db	149	TTAGTAGAGAACCTACACCTGTCAACTTAATTGGAAGAAATCTGTGACTCAAGATTGG-TGC	91	
Qy	661	ACCCTGAACCTTCGCCATCAGCCCATCGAACCGTGCCCGTGAAGCTGAAGCCCGGCAATG	720	
Db	90	ACCTTAATTTTCCCATTAGTCTCTATTGAAACGTACACGTAATAATTAAAGCAGGAATG	31	
Qy	721	GACGGCCCCCAAGGTAG	738	
Db	30	GATGGCCCAAAAGTAAAG	13	

RESULT 2	CL974989	LOCUS	DEFINITION
	CL974989	2011 bp	DNA linear GSS 21-SEP-2004
	OSIFCC042895		<i>Oryza sativa</i> Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

ACCESSION	CL974989
VERSION	CL974989.1
KEYWORDS	GI:52404497
SOURCE	GS.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 (bases 1 to 2031)
 Ma, L., Manoc, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 REFERENCE
 AUTHORS

Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G.K.S., Deng, X.W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
Unpublished (2004)

COMMENT: Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559

Tel.: 86-10-80488572
 Fax: 86-10-80488576
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Download/Out: 1454000

```

1. 2031
   Location/Qualifiers
     organism="Oryza sativa (indica cultivar-group)"
     mol_type="genomic DNA"
     db_xref="taxon:39946"

```

Query Match	Score	DB	Length	Origins
4.0%	97.8	10	2031	/clone_id=Oryza sativa express library" /note="Oryza sativa exon trapped genomic sequences "

Best Local Similarity 42.6%; Pred. No. 2.7e-07;
Matches 787; Conservative 0; Mismatches 1032; Indels 30; Gaps 4

153 CGGCACGACGTACTCGTCGCTCGAGGTCTACCGAAGCGGCATGTCGAGATCATATGGCCA 212

QY CAACATGCTGACCCGAGCTGGGCTGTGACCCCTGAATCTTCCCATACAGCCCATCGAACCGT 695

Db CGACCGAGGGAACTCGGATTCACGCCCTCGTGGGTGCGCTTTCACCGACGGCGGCGGGCT 272

QY GCCCGTGAAGCTGAAGCCCGGCATGACGAGCCCGCAAGGTGAAGCAGTGGGCCCTGACCGA 755

Db CATCGGTGAGGCCGCGCAAGAACGAGGCGGCGCAACCCGGAGCGACATCTACGACGC 332

QY GGAAGAAATCAAGGCCCTTGACCGGCATTTGGGAGGAGATGGAAGAAAGAGGCAAAATCAC 815

Db CAAGCGGCTCATCGGCGCGGAGTTTCTCGAAGCGCCGAGGTGACGCGCATGAAAGCTGCT 392

QY CAAATCGGCCCCCGAAGACCCCTTCAACAACCCCGTGTTCGCATTCAGAAAGAAAGACAG 875

Db GCCGTTCGCCCTGTCGACCGGAACGGGCAAGCGGCACGTCGCGTCGAGGTGAAGAGACGG 452

QY CACCAAGTGGGCAAGCTGTGTGACTTCGCGGAGCTGAACAAGGCGACCCAGGACTTCTTG 935

Db CGACGTGTG-----CGGTTCAGCCCGGAGGAGGTGAACGCGCATGTGTCTCAC 500

QY GAGGGTGAGCTGGGGCATCCCCACCCGCGCGCTGAAGAAAGAAAGAGCGGTACCGT 995

Db GCGATGAAGAGACGCGCCGAGGCTTACTTGGCGAAGAGGTCAACCCGCGCGCTGTAC 560

QY GCTGGAACGTGGGCAACGCTTACTTTCAGCGTGCCTCTGACGAGGACTTCCGCAAGTACAC 1051

Db CGTCCCGGCTCATTTCAACGACGCGGACGGCGAGGCCAACGAAGACGCGCGGTATCGC 620

Oy 1056 GCCTTCACCATCCCCAGCATCAACAGAGACCCCGGAGTTCGGCTACAGTACACGT 1111
 |||
 Db 621 CGGGTTCACCGTCGACCGCATCATCAACGAGGCCACCGCGCGGCATCGCTTACGGGAT 680

0y	1116	GCTGCCCGAGGGCGAGCCCGAGATCTTCCAGAGACACATGACCAAACTCT	1172
0y	681	CGACGAGAGGGCCCGGAGGAAGATGTCCTGCTTCCACTCTGGCGGCGGCGACCTTTCCA	740
Db			
0y	1116	GGAGCGCTTCGCGCGCCGCAACCCCGAGATCGTGAATCTACACGATCATGAGACACTGTA	1231

Db 741 CATTAGCATCTCGCATTCACAAACGCGCTGTTCAAGSTCTTTCACCAACGCGACAC 800
Oy 1236 CATTGCGACGACCTTGAGATCGGCCACACCGCGCCAAAGATCGAGGACTCGCAGACA 1291

Dd 801 CCACTTCGCGCAGAGACTTGGACAAACGCTCATGAGCCATTGCATGAAGTATCCG 860

Oy 1296 CCTGTGCGCTGCGGCTTCACCACCCCCGACAGAAGAACCAAGAGAGGCCCTTTCTT 1351

Dh 861 CTGGAAGATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 920

1356 GTGAGTGGGCTACGAGCTGCACCCCGACAAGTGAACCGTGACGCCATCGAGCT---GCC 1411
 921 GTGCGAGGCGGCCCAAGCGCGCGCTTCAGGACCAAGACCAAGTGGCGCGTGCAGGTGAGTGC 980

Accession	Sequence	Length
Oy	CGAAGAGAGAGCTGGACCGTGAACACATCCAGAGCTGTGTGGCAAGCTGAACCTGGGC	147
Db	981 CCGTTCGACGAGCTTGTCGGAACCGCTTCCGGGGCGCGTTCAAGAGCTCAA	104

153
1473 CAGCGAGATACCCCCGACATCAAGATGCGCCAGGTGAGCAAGCTTCATTCGCGCGGCCAA 153
DB 1041 CAGCGACCTCTTCAAGAAGACGATGTGTCCGGGTAGAAAGCCCATGCGCGAGCGCCCGGCT 11041
QY 1553 GAGCCTTACCGACATCGTGGCCCTGATCCGAGAGAGGCCGAGGCTTGAGAGCTTGCCCGAAGACCG 159

Db 1101 GAGCAGAGGCGACATCGACGAGATCGTCTCGTCGGCGGACACGAGATCCCAAGT 116
QY 1593 CGAGATCTGCGCCGAGCCCGTCACAGCGGTGTACTACACCCAGACGAGACCTGGTGC 165

Accession	Sequence	Length
Db	1161 GCAGCAGCTGCTCAAGGACTTACTTCCTGGTGGCAAGAGGCCCAACCCGCGGCTCAACCCCGA	122
Dy	1563 CGAGATCCAGAAAGCAGGGGCGCAACACCAAGTGAACCTTACCAAGATCTTACCAAGAGCCCTTTCA	171
Db	1221 CGAGGCCCTGCGCTACGGCGCGCGCGCTCAGAGCCAGCATTCATCACCGGCGCAGTGGACCA	128

QY 1713 GAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCGCCCAACCAACGAGCTGAAGCA 1772
 |||||
 DB 1281 GAAACACGAGGACATGATCTCTCGACGTCGCGCTCACCTTCGCGCTGAGACGCGC 1340
 QY 1773 GGTGACCGAGGCGCGTCAGAAAGATGCGCATTGAGAGACATGATCTGGGGCAAGACCCC 1832
 |||||
 DB 1341 CGGCGCGGTGATGCGCCAACTGATCCCGGCAACACGCTGTCGACGACGAGAGAGCGCA 1400
 QY 1833 CAAGTTCGCGCTGCCCATCCAGAGAGACCTTGAGAGACCTGCTGAGACCGACTACTGAGCA 1892
 |||||
 DB 1401 GGTGTTTCAACACGATCAAGAGACAGAGACACACGTCATCATCAGAGTGTTCAGAGGCGCA 1460
 QY 1893 GGCACCTTGATCCCGAGTGGAGTTCTGTAACACCCCCCTTGTAAGCTGTGTA 1952
 |||||
 DB 1461 GCGGACATGACGAGGAGCAACCGGCTGCTCGCAGATTCGACCTTCGCGCATCCGCGC 1520
 QY 1953 CCAAGCTGAGAGAGAGGCCATCATGCGCGCCGACCTTCTCATGTCGTAAGCGCGCGCCCA 2012
 |||||
 DB 1521 GCGCGCGAGGCGCGCGCGCGAGATCGAGGTGACG-----TTGAGGTGAGCGCGCA 1571
 QY 2013 CCGGAGACCAAGATCGGCAAGGCGGCTACGTCAGACCGAGGCGCGCAAGAGATCGT 2072
 |||||
 DB 1572 CGGCACTCTCAAGCTCTGCGCGCGCGCAAGGCGCACCGGAGGTTCGAGAGATCACT 1631
 QY 2073 GAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAAGGCGCATCAGCTGCGCTGCA 2132
 |||||
 DB 1632 CTCGCGGACGACCGCAAGATCGCAGAGAGATGACCGGATGATGTCGCGAGCGGCA 1691
 QY 2133 GGAACGCGGACGAGAGGTGAACCTGTAACCGACACGCTGTAAGCCCTGGGCTATCCA 2192
 |||||
 DB 1692 GGAAGTTGCGCGAGAGAGACCGCGGCAACAGGAGAGGTGAGACGCTCGGAAACGCTGGA 1751
 QY 2193 GCGCCGCGCGCAAGAGCGAGCGAGCTGTGAACCAAGT-----CATCGAGACGT 2246
 |||||
 DB 1752 GCGCTACGCTTACATCAATCAAGACACGCTCGCGCGCAAGATGCGGACCGCATGAGAG 1811
 QY 2247 GATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGCGCGCGCAAGAGGCGCATCGCGGCA 2306
 |||||
 DB 1812 CGAGGAGAGAGACAGAGTGAAGAGCGGCTGAGAGGCGGTACGAGTGGCTGGAACGCGCA 1871
 QY 2307 CGAGCAGATCGACACACTGTGTGAGCAAGGCGCATCCGCAAGTGTCTGTCGACGCGAT 2366
 |||||
 DB 1872 CCGGAGCGCGCGCAAGAGAGTACGAGAGAGCTGAGAGACTGAGAGACCTGTGCA 1931
 QY 2367 CGATGCGGCGATGATCTACCAAGTACATGACGACCTGTACGTGAGC 2415
 |||||
 DB 1932 CCGCTCATGTGCGCGGTCTACCAAGAGTCCGCGGCGCGCGCGCGC 1980

RESULT 3
 CL967755 2886 bp DNA linear GSS 21-SEP-2004
 LOCUS OsIRFCC015718 Oryza sativa Expressed Library Oryza sativa (indica
 DEFINITION cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL967755
 VERSION CL967755.1 GI:52390149
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 BUKARYOTA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriaristideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 2886)
 AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W. and Wang, J.
 TITLE An analysis of transcriptional regulation of the rice genome and
 JOURNAL its comparison to Arabidopsis
 COMMENT Unpublished (2004)
 CONTACT Chen Chen
 DEPARTMENT Department of Bioinformatic
 BEIJING INSTITUTE OF GENOMICS

Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.

FEATURES

source

1..2886
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 3.9%; Score 97.2; DB 10; Length 2886;
 Best Local Similarity 43.2%; Pred. No. 3.5e-07;
 Matches 712; Conservative 0; Mismatches 903; Indels 33; Gaps 4;

QY 189 GATGAGAGCTGACAGCGGCAAGGCGCAACTTCTTCGCGAGAGACTTGCTTCCCA 248
 |||||
 DB 219 GGTGATGAGGCGCGCGCCATCATGCTGCTCGCCAAACGCGCGCGCGCGCGC 278
 QY 249 GGGCAAGCGCGGAGTTCCGAGGAGCAACCGCGCCACAGCCCGACCGCGCA 308
 |||||
 DB 279 CCACTGGACGAGACTTCTGTCGTCGTCCTCTCTGATCATCTCACTCATGTTT 338
 QY 309 GCTGCAAGTGGCC---GGGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCA 365
 |||||
 DB 339 CATGAGAGAGAGATGCGCGGAGAGCGCGCGCGCGCTCATGAGCGCGCTGCGCGCA 398
 QY 366 CCTGAACCTTCCCGAGATCACTTGTGGAGCGCGCGCGCTGTGAGATCAAGTGGCGG 425
 |||||
 DB 399 GACCAAGGTGCTCAGAGAGAGGAGAAATGGCAGAGAGAGAGAGCGGTCTCTGTCGCG 458
 QY 426 CCAAGATCAAGAGAGCGCTTGGACACGCGCGCGCAACACCGTCTGAGAGAGATGAG 485
 |||||
 DB 459 CCAATCATGACATCAAGCTGCGGACATCATCTCCGCGAGCGCGCTGTCAGAGG 518
 QY 486 CTTGCCCGCAAGTGAAGACCAAGATGATCGCGCATCGCGCTTATCAAGGTGCG 545
 |||||
 DB 519 CACCGCTCTAAGTGCACAGCGCGCGCTCACCGCGAGTGCATCCGTCAACAGCA 578
 QY 546 CCAAGTCAACAGATCTGATGAGATCTGCGGCAAGAGGCGCATCGGCACTGTCTGAT 605
 |||||
 DB 579 CCGCGGCGAGGCGGTCTTCTCGGAGTCAACCGTCAAGAGGCGGAGATCGAGGCGGTG 638
 QY 606 CCGCCCAACCCCGTGAACATCATCGGCGGCAAGTGTGACCACTGGCTGCAACCT 665
 |||||
 DB 639 CATGCGCACCGGCTGACACCTTCTTGGCAAGGCGCGCACTTGTGTGACGACACCA 698
 QY 666 GAACTTCCCATATGAGCCCATGAGACGCTGCGCGGAGAGCTGAACCCCGCATGAGCG 725
 |||||
 DB 699 CAACATGCGGCACTTCACTGCTGTGCTCAACGCGCATTCGCAATCTTGATCATTCAT 758
 QY 726 CCGCAAGTGAAGCACTGCGCGCTGACCGAGAGAGATCAAGGCGCTTACCGCATCTG 785
 |||||
 DB 759 CCGCTGCGCATGATATGAGATCATCTGTATGTCACCGATCAGACCGCGCGTACCG 818
 QY 786 CCAAGAGATGAGAGAGAGGCAAGTCAACCAAGATCGGCGCGGAGAACCCCTCAACAC 845
 |||||
 DB 819 CAGCGCATCGACACCTGCTGCTGCTCATCGGCGGAGATCCCATCGCATGCCAC 878
 QY 846 CCGCTGTTGCGCATCAAGAGAGAGACAGACCAAGTGGCGGAGCTGTGATCTTCCG 905
 |||||
 DB 879 CTGTGCTGCTGATCATGAGCATGCGGTCCACCGGCTGTGACAGAGGCGCATAC 938
 QY 906 CAGCTGAACAGCGCACCCAGACTTCTGGAGGTGCACTGGGCACTCCCAACCCCGC 965
 |||||
 DB 939 CAAGCGATGACCGCATGAGAGATGCGCGGCAAGAGATGCTGTGACGACCAAGAC 998
 QY 966 CCGCTGAGAGAGAGAGAGCGGTGACCGTGTGAGCTGTGGGAGAGCGCTTACAGCGT 1025
 |||||

Db	999	CGGGAACGCTGAACGCTCAACAAAGACTCAACCTGTGACAAAGACGCTGATTCGAGGTGTACGGGGCG	1058
Oy	1026	GCCCTTGGACGAGGACTCTCCGCAAGTACACCGCTTCAACATCCCAAGCATCAACAAACGA	1085
Db	1059	GGGGCTGGACAAGGACTCGTGCTCTCTGTATGCGCGGACGAGGGCGTCTCCGCGTGAAGAACCA	1118
Oy	1086	GACCCCGGACATCCGACTACAGTACCAAGTCTGTGCTGGCCCAAGGGCTGTGAAGGGCAGCCCGAG	1145
Db	1119	GGAAGCCATTGACACATGATATCTGTGGGATGCTCGCCCAACCCCAAGAGAGCCCGCGCCGG	1178
Oy	1146	CATCTTCCAGAGCAGCATGACCAAGATCTTGAAGCCTTCCGCGCCCGCAACCCCGAGAT	1205
Db	1179	CAT-----CAAGGAGTTCACCTTCTCCGTTCAACCCCGTCCAGAGAGCCGCAAGGCG	1229
Oy	1206	CGTATCTTACCAAGTACATGAGCAACCTGTATGTTGGGACGACACTGTGAGATTCGGCCACGA	1265
Db	1230	CATCACTTACATCGACCGCAATGGGAAATGGCACAGATTCACGAAGGGCGCGCGAGACCA	1289
Oy	1266	CGCGCCCAAGATTCGAGGAGCTGCGCAACACTGTGCGCTGAGGGCTTCAACACCCCGCA	1325
Db	1290	GATCATCGAGCTGTGCAAGATGAGCAAGACGCGCAAGAAAGGTGTGCAACGCTGATCGA	1349
Oy	1326	CAAGAAAGACCAAGAAAGAGCCCTTCTGTGATGGGCTTACGAGCTGCAACCCCGCAAA	1385
Db	1350	CCAGTACCGGACCGCGGCTTCGCTGCTGGCGTGTCTGTACCAAGAAAGGTGCGCGAGAA	1409
Oy	1386	GTGACACGTGACGCCCATTCGAGCTGCTCCGCAAGAGAGAGACTGGAACCGTGAACGATCCA	1445
Db	1410	G-----AGCAAGAGAGAGAGAGGGGAGGCCATGCGAGTTCTGTGCGGCTTCT	1454
Oy	1446	GAAGCTGTGGGCAAGCTGAACTGGAGCAGCCAGATCTTACCCCGGACATCAAGGTCGCGCA	1505
Db	1455	GCCGCTGTTCGACCCGCGCAAGGACCAACAGCGCGGAGACCACTCCGCGCGCTGCACCT	1514
Oy	1506	GCTGTGCAAGCTGCTGCGCGGCGCAAGAGCCCTGACCGACATTCGTGCCCCCTGACCCGAGA	1555
Db	1515	CGGCGTGAACGTGAAGATGATTCACCGCGCAACGCTGCCCAATCG-----GAAAGAGAAC	1568
Oy	1566	GAGCCAGCTGAGGCTTGCGCGCAAGAACCGCGAGATCTTGCGCAAGCCCGTGCACCGCGGTGA	1625
Db	1569	GCGCGGCGGCGCTCGGAGATGGGACCAACATGTATCCGTGTGACGACGCTGTGGGGGACAA	1628
Oy	1626	CTACGACCCCAAGAGACCTGTGTGCGCGGATCTTCCACGAAGCAAGGACCAAGTGAAC	1685
Db	1629	GAGCAGCCAGATGAGACCGGCTTCCCAATGAGCGAGCTGTATCGAAGAGCGCGACGCTTCGC	1688
Oy	1686	CTACCAAGATCTTACCAAGAGCCCTTCAAGAACCTGAAAGCCGGCAAGTACCGCAAGATGCG	1745
Db	1689	CGGCTGTCTCCCGAGAGCAACAGTACAGATGTGTGAAGCGCTTCCAGAGCCGAAAGCACAT	1748
Oy	1746	CACCGCCCAACCAACGAGCTGTAAAGCACTGACCGAGGCGGTGCAGAAAGATCGCCATGGA	1805
Db	1749	CTGCGGAGCATGACCGGCGCAGCGCGGTGAACGAGCGCGCGCGCTGAAGAAAGCCGACATCGG	1808
Oy	1806	GAGCATTCGTGATCTTGGGGCAAGACCCCC	1833
Db	1809	CATGCGCGTTCGACGACGCGCACGAGAGCC	1836

Db 661 AAGACCTGACCGCTGCTCCAGTGCACGCCCGCGCTGGAAGACGTCGCTTCGTC 720

Qy 1150 TTCCAGAGCAGATGACCAAGATCTGAGAGCCCTTCGCGCCCGCAACCCGAGATGCTG 1209

Db 721 ACCTACCAACGACGACCAACGTCGCGCTCCGACGCGAGCTCGCTGCGTCTTC 780

Qy 1210 ATCTACGATACATGACGACCTGTACGTGGGAGGAGCTGAGATGCGCCAGACCCG 1269

Db 781 TGGATGTTCATCGCGGAGAGCTCGCGCTGTGTCGCGCCGCGCTCGAGCGGCTCATC 840

Qy 1270 GCCAAGATGAGAGAGCTGGGAGCAGCAGCTGTGCGGTGGGCTTCACACCCCGACAG 1329

Db 841 CTGTGAGACGGTGGCTCCGACCGCGCTGTCCGCTGATTTCCGACACGAGGTCAAG 900

Qy 1330 A 1330

Db 901 A 901

RESULT 5
AY103647 2598 bp mRNA 1linear HTC 18-FEB-2005

LOCUS Zea mays PC0142084 mRNA sequence.
ACCESSION AY103647
VERSION AY103647.1 GI:21206725
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Bukariyota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogonaceae; Zea.
1 (bases 1 to 2598)
Gardiner,J., Schroeder,S., Polacco,M.L., Sanchez-Villada,H., Fang,Z., Morgan,M., Landewe,T., Fessler,K., Uesche,F., Hanfey,M., Tingey,S., Chou,H., Wang,R., Soderlund,C. and Coe,B.H. Jr.
Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional plant Physiol. 134 (4), 1317-1326 (2004)

JOURNAL PUBMED 15020742
REFERENCE 2 (bases 1 to 2598)
AUTHORS Hailey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsett,M.S., Arthur,L.W., Hanfey,M., Morgan,M. and Tingey,S.V.
TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 3 (bases 1 to 2598)
AUTHORS Coe,B.H.
TITLE Direct Submision
COMMENT Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at NSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES
source
1..2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/Dupont Consensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize

ORIGIN Mapping Project"

Query Match 3.7%; Score 90.8; DB 4; Length 2598;
Best Local Similarity 43.3%; Pred. No. 4.9e-06;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

Qy 170 GCGGCAAGAGGCGCCACGATGAGACCTGACCCGAGGCGCAGGCACTTCTCCGG 229

Db 351 GAGGAGAGTGAATCTGACTGATCGCCACGCGGCGGACATCTACGCGTCAACACCG 410

Qy 230 AGACCTGACCTTCCCGGAGGCGGAGGCGGAGGCTTCCGAGGAGGAGGCGGCGCA 289

Db 411 GCTTGGGCGGACCTTCCACCGCGGACCAAGAGCGGCGCGCTCCAGTGAAGTCC 470

Qy 290 ACAGCCCAACGAGCGGAGCTGAGGTGCGGCGCAACCCCGCAGCGAGGCGGCG 349

Db 471 TCAGGCAATCTCAACCGCGGATCTTTCGACCGGCGAGCGGCGACGCTGCGG 530

Qy 350 CCGAGCGCGAGGCAACCTGAACTTCCCGGATCAACCTGTGGGAGGCGCGCGTGTGA 409

Db 531 AGGTACCGCGCGGAGTGTGTGCGATCAACCTCTCTCCAGGCTACTCCGCGCA 590

Qy 410 GCATCAAGGTGGGCGGCGGAGTCAAGAGGCGCTGTGACACCGGCGCGACGACCG 469

Db 591 TCCGCTTGAGATCTTCAGAGGCTCAAGAGCTGTCAACACGCGTTCAGCCCTGCC 650

Qy 470 TGTGAGAGAGATGAGCTCCCGGAGTGAAGCGCAAGATGATCGGCGGATCGCG 529

Db 651 TCCGCTCGGCGGCGACATCAACCGGCGGCGGAGCTGTCCCGCTCTCAATCGCG 710

Qy 530 GCTTCATCAAGTGGCGGCGGAGTCAAGATCTGATCGAGATCTGCGGCAAGAGCGCA 589

Db 711 GCTTATCAACGCGCGCGCGCGCAAGCGGAGCGGCTACCG---TCGACGGAAGAGGTGG 767

Qy 590 TCGGACCGGTGATCGGCGCGCGCGCGCGGAGATCAATCGGCGGCAATGTGACCC 649

Db 768 AGCGCGCGGAGGCGGCTTCAAGATCGCGGATGAGGCGGCTTTCAGAGTCAACCGCA 827

Qy 650 AGCTGGGCTGACCCCTGAATCTTCCCATGAGCGGCGGATCGAGCCGCTGAGAGCTGA 709

Db 828 AGGAGGCGCTCGCATCGTCAACGAGCGAGCTCGTGGGCTCCGCGCTCGGCGCACGTTGA 887

Qy 710 AGCGCGGATGAGCGGCGCGGAGTGAAGATGAGCGGCGGCTGACCGGAGGAGATCAAG 769

Db 888 TGTACGACCGCAACGCTCTGCGCGCTCTCTGCGAGGTCTGTCTCGCGCTTCTGTGAGG 947

Qy 770 CCTGACCGCATCTGCGGAGATGAGAGAGGCGGAGATCAACAGATCGGCGCG 829

Db 948 TATGAGCGGCAAGCGCGGATCAAGAGCACTGACCCAGAGCTGAGAGCACACCGG 1007

Qy 830 AGAACCTTACCAACCGCGGCTGTTCGATCAAGAGAGGAGGAGGAGGAGGAGGAGGAG 889

Db 1008 GGTTCATGAGGCGCGGCGGATCATATGAGACATCTGTGATGAGGAGGAGGAGGAGGAG 1067

Qy 890 AGCTGTGACTTCCGCGGAGTGAACAGCGGACCGAGACTTCTGGAGGTGAGCTGG 949

Db 1068 AGCGCAAGAGTGAACGAGCTGAGCGCGC-----TGTGAAGCGGAGGAGGAGGAG 1115

Qy 950 GCATCCCGCAACCGCGCGGCTGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1009

Db 1116 ACAGGTACGCGCTCGCGAGTGCAGGAGTGTGGGCGCGCGGAGTGAAGTATCGCG 1175

Qy 1010 AGCGCTTACTGAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1069

Db 1176 CGGCAACGAGTCAATGAGGAGGAGGAGTCACTCGGTGAACAGCAACCGGATCATGAGG 1235

Qy 1070 CCAAGATCAACGAGAGCGCGGAGTCCGCTACGATCAACAGTGTGCGCGAGGCT 1129

Db 1236 TCACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1295

Qy 1130 GGAAGGAGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1189

Db 1236 TGGACACGCCCCCTCGCCATCGCCACATCGGCAAGCTCATGTTCCGCGACTTCTCCG 1355
 QY 1130 CCGGACACCGGAGATCGATCTACCACTACATGAGACGCTGTACGCGGACGAG-- 1247
 Db 1336 AGCTGTACACGAGTTCTACCAACACGCGCTCACTTCAACTGGCGGACCGCAACC 1415
 QY 1248 ---CCTGAGATCGCGCAGCAACCGCGCCAGATCGAGAGCTGCGCAAGCACTGTCC 1303
 Db 1416 CCAAGCTGTACACCGCTTCAAGGGGACCGAGATGCCATGGCCTCTACTGTCTCCAGC 1475
 QY 1304 GCTGGGGCTTCAACACACCCCGACAGAGACCAAGAGAGGCCCTTCTGTGTAGTG 1363
 Db 1476 TCACATGCTTGGGCAACCCCATCACCAACATGTGAGCGCGGAGACCAACAG 1535
 QY 1364 GGTACAGCTGACCCCGACAGTGAACCGTGAAGCCCATGAGCTGCCGAGAGAGAGA 1423
 Db 1536 AGTGAATCTCTTGGGCTCTGTCTGCGCCAGAGAAAGCCCGGATGTACATCTTA 1595
 QY 1424 GCTGACCGGTGAACGACATCGAAGAGCTGTGGCAAGCTGAATGGGCAAGCAAGTCT 1483
 Db 1596 AGCTGATGTCTCACTACATCATGTGGGCTGTGGCAGCGCGTGGACCTGGCGCACTCG 1655
 QY 1484 ACCCGGACATCAAGTGTGCGCCAGCTGTGCAAGCTGTGCGCGCGCCAGAGCCCTGACCG 1543
 Db 1656 AGAGAAACATCAAGGCTGTGTGAGAAACACCGTGAACCGAGTGGCCAGAGAGTGTGA 1715
 QY 1544 ACATCGTGTCCCTGACCGAGAGAGCGCGAGCTGGAAGTGGCGAGAA-----CGCGC 1594
 Db 1716 CATGAACCTCTCGGGCGAGCTCTCAAGCGCCGCTTCAAGCAAGAGAGCTGATAGG 1775
 QY 1595 AGATCTGTGCGAGCGCGTGTGACAGCGGTACTACAGACCCAGCAAGACCTGTGTGCGCG 1654
 Db 1776 CATGACACCGGAGCGCGTGTGTACAGTACGCGAGAGACCGCGGCGAGCGCTGCGCG 1835
 QY 1655 AGATCAAGAGAGAGCGCGCAGACAGCACTGAGACTTACCAATCTTACAGAGAGCCCTTCAAGA 1714
 Db 1836 TGATCGAAGAGTGTGCGCGCGTGTGTGTGACCAACCCCTTCAAGACGCG---GCGAGCGCG 1892
 QY 1715 ACCTGAAGACCGGCAAGTACGCGCAAGATGCGACCGCGCCAGCAACAGAGTGAAGCGCG 1774
 Db 1893 AGCGGAGCGCTTCCGTGTCTCAAGATACCAAGGTTCAAGAGAGAGCTTCCGCGCGGTGC 1952
 QY 1775 TGACGAGAGCGGTGAGAGATGCGCATGAGAGACATCTGTG 1815
 Db 1953 TCCCCAGAGAGGTGAGAGCGCGCGCGTGTGGGTGCGCGAG 1993
 RESULT 6
 LOCUS C2247380 743 bp DNA linear GSS 11-FEB-2005
 DEFINITION A1A4-asf35a12.b1 Ancylostoma caninum whole genome shotgun library
 (A1A4GSS 001) Ancylostoma caninum genomic, genomic survey sequence.
 ACCESSION C2247380
 VERSION C2247380.1 GI:59632821
 KEYWORDS GSS.
 SOURCE Ancylostoma caninum (dog hookworm)
 ORGANISM Ancylostoma caninum
 Bacteria; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 REFERENCE 1 (bases 1 to 743)
 Miltreva, M., McCarter, J.P., Pape, D., Rittler, R., Tsagaris, H., R.,
 Ronko, I., Martin, J., Mylie, T., Dante, M., Meyer, R., Messing, D.,
 Waterson, R.H., Clifton, S.W. and Wilson, R.
 Genome Survey sequences from the parasitic nematode Ancylostoma
 caninum
 Unpublished (2004)
 CONTACT: Miltreva M
 Washington University in St. Louis
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: nematode@watson.wustl.edu

FEATURES
 source
 location/Qualifiers
 1..743
 /organism="Ancylostoma caninum"
 /mol_type="genomic DNA"
 /strain="Baltimore"
 /db_xref="taxon:29170"
 /dev_stage="Adult"
 /lab_host="GS10"
 /clone_1ib="Ancylostoma caninum whole genome shotgun
 library (A1A4GSS 001)"
 /note="Vector: pOTM13; Site 1: BstXI; Site 2: BstXI;
 Ancylostoma caninum genomic DNA was randomly sheared,
 end-repaired and size fractionated to enrich for 2-4 kb
 fragments. Genomic DNA was provided by John Hawdon
 (mtmj@wumc.edu) at George Washington University.
 Sequencing by Washington University Genome Sequencing
 Center, St. Louis, MO."

ORIGIN
 Query Match 3.6%; Score 89.4; DB 10; Length 743;
 Best Local Similarity 46.3%; Pred. No. 8e-06;
 Matches 294; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

751 ACCGAGGAGGATCAAGGCCCTGACCGCCATCTGCGAGAGATGAGAGAGGCGCAAG 810
 Db 63 AACGACACACACACACACACACACACACACACACACACACACACACACACACAC 122
 QY 811 ATACCAAGATCGCGCCGAGAACCCCTTACACACACCCCGTGTGCGCATCAAGAGAG 870
 Db 123 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 182
 QY 871 GACAGACACACACACACACACACACACACACACACACACACACACACACACAC 930
 Db 183 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 242
 QY 931 TTCTGGAGGTGCACTGTGGCATTCCTCCACCGCGCGCGCTGAGAGAGAGAGCGTG 990
 Db 243 GACACACACACACACACACACACACACACACACACACACACACACACACACAC 302
 QY 991 ACCGTGTGACAGTGTGCGACGCGCTTACATTCAGCGTGCCTTGAAGAGAGACTTCCGCAAG 1050
 Db 303 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 362
 QY 1051 TAGACCGCTTACCATCCCGACATCAACACAGAGAGAGAGAGAGAGAGAGAGAG 1110
 Db 363 AACGACACACACACACACACACACACACACACACACACACACACACACACACAC 422
 QY 1111 AACGTGTGCGCCGAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
 Db 423 AACGACACACACACACACACACACACACACACACACACACACACACACACACAC 482
 QY 1171 ATCTGTGAGGCTTCCGCGCGCGACACCCGAGATGTGTATCTACAGTACATGAGACAC 1230
 Db 483 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 542
 QY 1231 CTGTAGTGTGCGAGGAGCTGTGAGATCGCGCAGACCGCGCCAGATGAGAGAGTGTGCGC 1290
 Db 543 AACGACACACACACACACACACACACACACACACACACACACACACACACACAC 602
 QY 1291 AAGGACCTGTGCGGTGTGAGGCTTCAACACCCCGACAGAGAGAGAGAGAGAGAGAG 1350
 Db 603 AACGACACACACACACACACACACACACACACACACACACACACACACACACAC 662
 QY 1351 TTCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1385
 Db 663 TACACACACACACACACACACACACACACACACACACACACACACACACACAC 697

RESULT 7
 C1961989

LOCUS CL961989 1398 bp DNA linear GSS 21-SEP-2004
 DEFINITION O81PCC006991 *Oryza sativa* Express Library *Oryza sativa* (indica
 cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL961989
 VERSION CL961989.1 GI:52378720
 KEYWORDS GSS.
 SOURCE
 ORGANISM *Oryza sativa* (indica cultivar-group)
Oryza sativa (indica cultivar-group)
 Burkayota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; *Oryza*.
 1 (bases 1 to 1398)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W. and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 CONTACT Chen Chen
 Department of Bioinformatics
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Location/Qualifiers
 1..1398
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_1b="Oryza sativa Express Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN
 Query Match 3.5%; Score 86.6; DB 10; Length 1398;
 Best Local Similarity 45.9%; Pred. No. 2.7e-05;
 Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;

```

Oy 1  GTGACGCGCAATGCGCGGCGCATGAGCGACGCGCGCAACGCGCATTCGATGCG 60
Db 166 GTGAGAGGACGTCGCGGCGGCGACATGCTTGTCTCGTGAACCGCCCGACCAAG 225
Oy 61  CCGAGCACTTCAGAGGCGCCCGCATCATCAAGTGTTCMACTGCGCGCAAGAGGCG 120
Db 226 GCCCGGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 282
Oy 121 CACATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db 283 CGGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 342
Oy 181 GCGCAACCAATGAAAGATGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db 343 GTCAACACCGCGGAGCGCATGCAAGATCTCTGACCAACCGCGCGCGCGCGCGCG 402
Oy 241 TTCGCCAGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db 403 TTCGAATCTCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 462
Oy 301 AGCGGAGCTGCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db 463 GCGCCCGCA-----CGCGCTCTCATCGCGCGCGCGCGCGCGCGCGCGCGCG 513
Oy 361 GGAACCTGAATCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db 514 GCGGTCAAGCGCTCAAGAGCGTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 570
Oy 421 GCGCGCAAGTCAAGAGCGCGCTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCG 480
Db 571 ACCACCAACTCTGTGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
Oy 481 ATGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540

```

```

Db 631 AGGATCTGCTGCTTCAAGCGCATGTCCGCGCTCTGCGAGGCGCACCGCGCGCGCGCG 690
Oy 541 GTGCCCGAGTACGACAGATCTGATGCA---GATCTGCGGAGAAAGCGCATCGGACG 597
Db 691 GAGGTGCGCTTACGCGGTGGCGAAGGATCTCAAGATCTGCGCGCGCGCGCGCGCGCG 750
Oy 598 GTGCTGATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 657
Db 751 GTGGGATTCGAGGCTCTGCTTCCAGAAAGACATCTCTCAACCTGCG----- 798
Oy 658 TGACCCCTGAATCTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 717
Db 799 TACATCTGCGAGTGAATGAGCTCTCCCGAGGTGCGCACTATGAGAGGATTAACAG 858
Oy 718 ATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 777
Db 859 ATCAACGACTACAGAAAGCGGTTGCGAAGCGCGGTGTGTCATGTTCAACACG 918
Oy 778 GCCATGTGCGAGAGATGAGAGAGGCGCAAGATCAACAAATGCGCGCGCGCGCGCG 837
Db 919 GTGCGCGCGAGAAAGATCGCGTGTGCGCTTCCAGAAAGACACCGCGCGCACG 978
Oy 838 TACAACACCGCGGTGTGCGCGCATCAAGAGAGACAGCAAGTGTGCGCGCGCGCG 897
Db 979 AGGAGAGCGCGCGCATGACCTGTGCAAGGCGCTGATGTGCGCAAGGCGCAAGGTAG 1038
Oy 898 GACTTCGCGAGCTGAACAGCGCAACCGAGACTTCTGAGAGTGCAGCTGCGCATCCCC 957
Db 1039 ATCTACGACCGCGCGAGTGAAGAGAGACAGATGTCAGCGCGCGCGCGCGCGCG 1098
Oy 958 CACCGCGCGCGCTTGAAGAAAGAGAGCGTACCGGTGCTGAGCGTGTGCGCGCG 1017
Db 1099 GATTGAGATCAACCGGTTCATCTCCAGCGAGAGCGCGCGCGCGCGCGCGCGCG 1158
Oy 1018 TTCAGGTGCGCGCTGAGAGAGACTTCGCGAGTACACCGCGCTTCACATCCCGACATC 1077
Db 1159 GTGCGCTGAGAGCGCTTACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1218
Oy 1078 AACCAAGAGACCGCGCGCATCGCTAC 1104
Db 1219 TGGAGAGATTCAAGAGCGCTCGACTAC 1245

```

RESULT 8
 BM321451
 LOCUS BM321451
 DEFINITION 951 bp mRNA linear EST 03-JAN-2002
 rockefeller.0.1211 *Mastigamoeba balamuthi* lambda ZAP II library
Mastigamoeba balamuthi cDNA similar to adenovylhomocysteinease (BC
 3.3.1.1), mRNA sequence.
 ACCESSION BM321451
 VERSION BM321451.1 GI:1805857
 KEYWORDS EST.
 SOURCE *Mastigamoeba balamuthi*
Mastigamoeba balamuthi
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
 1 (bases 1 to 951)
 Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,
 Gordon, P., Durruffel, L., Gaasterland, T., Lopez, P., Muller, M. and
 Philippe, H.
 The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockefeller.edu
 Insert Length: 951 Std Error: 0.00
 POLYA=No.
 Location/Qualifiers
 1..951
 source


```

/organism="Maestigamoeba balamuthi"
/mol_type="rRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Maestigamoeba balamuthi lambda ZAP II library"
/notes="syn: Phreatamoeba balamuthi"

```

ORIGIN

Query Match 3.5%; Score 85.6; DB 3; Length 951;
 Best Local Similarity 46.0%; Pred. No. 3.9e-05;
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

```

Oy 1734 GCCCAAGATGCGCACCGCCACACCAACGCTGAGAGAGCTGACCGAGCGCGTGCAGAA 1793
Db 219 GCCCAAGATGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 278
Oy 1794 GATGCCCATGAGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1853
Db 279 CATCCGCGAGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 338
Oy 1854 GAAAGAGACTGTGGAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1913
Db 339 GAGAGTGCACCTGTGAAAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 398
Oy 1914 GAGAGTGTGTGAAACACCCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1973
Db 399 CGAGCGCGGTGTACGCGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 458
Oy 1974 CATCGCGCGGAGACTTCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2033
Db 459 GCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 518
Oy 2034 GCGCGGCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2093
Db 519 GGTCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 578
Oy 2094 CCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2153
Db 579 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 638
Oy 2154 CATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2213
Db 639 GTTCCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 695
Oy 2214 GAGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2273
Db 696 CGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 755
Oy 2274 GTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2333
Db 756 CAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 815
Oy 2334 GGGCATTCGCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2393
Db 816 CGGCGAGGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 875
Oy 2394 CATGAGACACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2453
Db 876 GATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 917

```

RESULT 9
 CK159167/c
 LOCUS CK159167 869 bp mRNA linear BST 05-DEC-2003
 DEFINITION FGAS040564 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,
 mRNA sequence.
 ACCESSION CK159167
 VERSION CK159167.1 GI:38985053
 KEYWORDS EST
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE
 1 (bases 1 to 869)
 AUTHORS Allard, F., Crosby, W. L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroche, A.,
 Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D.,
 Peniket, C., Roach, J. L., and Sarhan, F.
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops
 Unpublished (2003)
 TITLE JOURNAL
 CONTACT Contact: Wm L. Crosby
 COMMENT Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgen_estecs.usask.ca

This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region [128,636].
 Plate: Talt537 row: N column: 23.

FEATURES

source

1..869

Location/Qualifiers

```

/organism="Triticum aestivum"
/mol_type="rRNA"
/cultivar="wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DH5 alpha"
/clone_lib="Triticum aestivum FGAS: Talt5"
/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression  

  subtractive hybridization) cDNA library from genotype  

  PI178383 cold hardened at 2 C for 21 days and 49 days  

  (equal amount of cDNA pooled together before subtraction,  

  tester) and subtracted against genotype Norstar cold  

  hardened at 2 C for 1 day (24 H) (driver). Modified Smart  

  cDNA (Clontech) priming and non-directional cloning"

```

ORIGIN

Query Match 3.4%; Score 85; DB 7; Length 869;
 Best Local Similarity 45.4%; Pred. No. 5e-05;
 Matches 304; Conservative 0; Mismatches 365; Indels 0; Gaps 0;

```

Oy 580 AAGAGGCGATCGGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 639
Db 827 ATGGCGGGCGCGCGACAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 768
Oy 640 ATGTGACCCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 699
Db 767 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 708
Oy 700 GTGAAGCTGAAGCCCGGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 759
Db 707 ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 648
Oy 760 AAGATCAAGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 819
Db 647 AACACACAGACACACACACACACACACACACACACACACACACACACACACACACAC 588
Oy 820 ATCGCGCGCGAGAACCTTCAACAACCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 879
Db 587 AACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 528
Oy 880 AAGTGGCGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 939
Db 527 AACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 468
Oy 940 GTGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 999
Db 467 AACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 408
Oy 1000 GACGTGGCGAGCGCTTACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1059

```

Db	407	GACAAACAAGCAACAAGACAACAACAAGCAACAACAACAACAACA	348
Oy	1060	TTCAACCATCCCAAGATCAACACAGAGACCCCGGATCCGCTACAGTACAAGTCTGTG	1115
Db	347	GACCAACAACGACAACGACAACAACAACAACAACAACAACAACAACAACA	288
Oy	1120	CCCAAGGCTGGAAGGGAGGCCGACATCTTCGAGACAGATGACCAAGATCTGTGAG	1175
Db	287	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	228
Oy	1180	CCCTTCGCGGCCCGCAACCCCGAGTCTGTATCTTACAGTACATGACGACCTGTACGTG	1235
Db	227	AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA	168
Oy	1240	GGCAGGCAC	1248
Db	167	AACGGCAAC	159

RESULT 10			
LOCUS	CL973991	3069 bp	DNA
DEFINITION	OS:FFCC025416 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.		

ACCESSION	CL973391
VERSION	CL973391.1
KEYWORDS	GI:52402507
SOURCE	GSS.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

REFERENCE
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
1 (bases 1 to 369)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriocaridaceae; Oryzaeae; Oryza.
1 (bases 1 to 369)

TITLE	An analysis of trancriptional regulation of the rice genome and its comparison to Arabidopsis unpublished (2004)
JOURNAL	
COMMENT	Contact : Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

```

FEATURES
  source
    location/Qualifiers
      1..3069
        /organism="Oryza sativa (indica cultivar-group)"
        /mol_type="genomic DNA"
        /db_xref="taxon:39946"
        /clone_lib="Oryza sativa Express Library"
        /note="Oryza sativa exon trapped genomic sequences "
ORIGIN

```

Query Match	3.4%	Score 84	DB 10	Length 3069
Best Local Similarity	42.8%	Pred. No. 8.2e-05		
Matches	536	Conservative	0	Mismatches 710, Indels 6, Gaps 2
QY	37	ACCACGCGCAATCTGTATGACGGCACAACCTTCAAGGCGCCCAAGGCATCATCAAG	96	
DB	625	ACCGGATATGGCAAGACGACGCTGGACACTACTCTTACCGAGAGACAGCATTCGGAAG	684	
QY	97	TGCTTCAACTGCGGGCAAGAGGGCCCATCTGCGCCCGCACTGCGCGGCCGCCCGCAAGAAG	156	
DB	685	CACCTTGCCTTGCACGGCGGTGGGGTCCACAGTCCCCGGAAGTTCCGCTTCCGCAACTCCCTTC	744	
QY	157	GGCTGTGGAAGTGCGGCAAGAGAGGGCCACAGATGAAGAATGCACTGCAACCGAGCGCCAGGCT	216	
DB	745	CACGACGTATCCGCGCAGGCCACAGCTCTTCCGCGGACAGACAGAGCTCGAGCGCGAGGC-	803	

QY	217	AACTTCTTTCGCGAGGACCTGAGCTTCTCCCGAGGGCAGAGGCCGCGAGTCTCCCAAGCAG	276
Db	804	--CGCCGACGACGACGTGCGGCGCCGCGGAGTGAATGAGGCGGAGAGGCTCTCAAGGCG	861
QY	277	CAGAACCGCGCCAACAGCCCAACGACCGCGAGACTGACAGTGCAGGCGGCGACMACCCCGCG	336
Db	862	CGCTTCGCGCGCGCGCTCCAGGGGAGAGATCACTGTGTGTCTGAAGCATCTCCGAGAC	921
QY	337	AGCGAGGCGCGCGCGGACCGCCAGGGACCCCTGAACTTCCCCCAATCAACCTGTGTGACG	396
Db	922	AGGAGGAGTGGGCTTCTTCTCTCGCGCGCTCCCGGCGGCTTCGCGGAGCTGCGTG	981
QY	397	CGCCCTCTGTGACGATCAAGGTGGCGGACGATCAAGAGGACCTGTGACACCGGCG	456
Db	982	CTCTGTCACGACGACGATCAAGATCAAGCGAGAGGTGAGGACMACCTGTGGCGCGCGCGGCG	104
QY	457	GCCGACGACACCGTCTGTGAGGAGATGAGACTTGCCTCCGGCAGTGGAACTCCCAAGATGATC	516
Db	1042	GCGCTGTGTGACCAACGAGGAGCTTCGAGAACTGCGCGGCGGAGACAGCGCCCGACGCTTTC	110

[illegible]

07 5111 GGCAGAAAGGCATGCAACGTCGTGATGAGCCGCCACCCGCGAACATCATGCGCC 658
 Db 1162 CTCGTGACCTCCATGACGAAAGGCGTCCAACTGCGCTCAACATGCTCATGCTCGCCGGA 122
 07 637 AACATCTGACCCAGCTGGGCTGCACCTTGAACTTTCCTCATGAGCCCATCGAGACCGTG 696

Db 1222 CTCCTGAGGTCGACGAAGAGGATGATGGAGCGCCGATGCATATGCTTCGATGATATAG 128

Qy 657 CCCGTGAAGCTGGAACCCCGGCATGACGCGCCCAAGAGTGAAACATGATGGCCCTTGACCGAG 756

Db 1282 CCACCCCGACATCCGCGAGCAGACGCGCGCGAGCGGACCGAGAGGACACACCTTC 134
 Qy 757 GAGGAAGATCAGAGCCCTTACCCGCAATCTGCGAGAGATGAGAGAGAGGCGCAAGATCAC 816
 Db 1342 GATGATTAGGAGACGCGCGCGCATTCAGACGAGCGAAACGAGAGAGCGAGAAAGAGGAAG 140

Accession	Sequence	Length
QY	AGATCGGCGCCCGAAGACCCCTTACACACACCCCGTGGTGGCCATCAAGAAAGAGACAC	876
Db	1402 CAGATGGGGGCACTTCATGCAATGCAAGAACTAGACAGGTGTGATGACACACTCTCCGACGCAT	146
QY	8777 ACCAAGTGGCGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGGCGCACCCAGGACTTCTGG	936
Db	1462 CTGAAGCCGTGTTCTCTTACTTCGCGCGGTTACCGCGAACAAGCCGATCTGGCGCGGG	152

QY 937 GAGGTCGACGTGGGATATCCCAACCCCGCGGCTTGATACAAAGAGGACGGACCG 956
Dlb 1522 AAGCTGGATCCGGCTGGGATCCCAAGAGGGTTTCAGACAGCCAAAGACGACGAGG 1588
QY 997 CTGACGTCGAGGACGACGCTACTTCAGCGTGCCTCGAAGACGAGACCTTCGCGAATGACACC 1055
Dlb 1582 GAGGAGCAACGGGAGAGAGTGCCTCAAGAGGCTATCTCCGGAGCTCGTCAATGACGCTGTC 1644

Qy 1057 GCGTTCACCATCCCGAGCTCAACACGAGATCCCCGGCATCCGCTTACAGTACCAACGATG 1117
Db 1642 GAGACGAGACCGCGCGCGCGCGAGGCGTCCGCGCTCAGCATCAACAAACGGGCTCTGAC 1707
Qy 1117 CTGCCCAAGGAGCTGAAAGGAGACGCCCAAGCATCTTCCAA--GAGCAGCATGACCAAGATC 1177
Db 1702 TTCTGCTCAAGCGGAGAGCGCGCGACACCAACTTCTCTCACTGACAGCGCGCGCGCGGAGC 1767
Qy 1174 CTGGAGCCCTTCCGCGCGCCGCAACCCCGAGATCTGATCTTACCAATGATGAGCACTCG 1237
Db 1762 CTCTCCAAACGGGCGCGCGCGCGCTCTCGGCAACCTTACGACACCGACCTCTGCGC 1827
Qy 1234 TACGTGGGCGAGCATCTGAGATCTGGCAGCAACCGCGCCAAAGATCGAGAGC 1285
Db 1822 GTCTATGCTCTGAGGCGCGCCCAAGCTCTCAACCTCTCTGTGAGCATCTCCGGAGC 1873

RESULT 11
 CL971508 1941 bp DNA linear GSS 21-BP-2004
 LOCUS OsJFCC021485 Oryza sativa Express Library Oryza sativa (indica
 DEFINITION cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL971508
 VERSION CL971508.1 GI:52397596
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiales; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 1941)
 AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W. and Wang, J.
 TITLE An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 JOURNAL Unpublished (2004)
 COMMENT Contact: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Location/Qualifiers
 source
 1..1941
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Express Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 3.4%; Score 83.8; DB 10; Length 1941;
 Best Local Similarity 42.9%; Pred. No. 8.6e-05;
 Matches 769; Conservative 0; Mismatches 992; Indels 30; Gaps 6;

257 CCCGACGAGCAAGAACCGGCGCAACAGCCCAAGCCGAGCTGAGGTGCGGCGA 326
 165 CGCCGCCAAGAACAGGTGCGCATGACCAACCAACCGCTTGATGCAAGCGGT 224
 327 CAACCCCGGAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 386
 225 GATTGATGAGAGGTTCTCGGACCGGTGCGTGAAGATGACATGAAAGCTATGCG 284
 387 CTTGTGAGAGCGCCCTGATGATGATCAAGGTGAGGCGGCGGCGGCGGCGGCT 446
 285 GGTGTGCTGCGCCCGGCGGAGCAAGCGATGATCTGTCAGTACAGAGGCGAGAGAA 344
 447 GGAACACCGGCGGCGGAGCAACCGGTGAGAGATGAGCTGCGCGGCAAGTGAAGCC 506
 345 GAGGTTCGCGGCGGAGAGATCTGTCATGATGATGATGATGATGATGATGATGAT 404
 507 CAAGATGATGCGGCGGATGAGGCGCTTCAAGGTGCGGCGGCGGCGGCGGCGGCT 566
 405 GGCCTTACCTCGGCTGCTCATCAAGAACGCGGTGTCACCGTCCGCGCTTACTTCAAG 464
 567 CGAAGATCTGCGGCAAGAACGCGATCGGCAAGGTGCTGATCGGCGGCGGCGGCGGCT 626
 465 CTCGCAAGAGGCGGCGGCAAGAACGCGGCGTCAATCGCGGCGCTCAAGTATCGCAT 524
 627 CATCGGCGGCAAGATGCTGATCCAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCT 686
 525 CATCAACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 584
 687 CGAGACCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 746
 585 CGGCGAAGAACGCTGCTCATCTTGAAGCTCGGCGGCGGCGGCGGCGGCGGCGGCT 644

OY 747 CTTGACGAGAGAGATCAAGCGCTTGAACCGCATCTGCGAGAGATGAGAGAGAG 806
 DB 645 CACCATGAGAGAGGAGATCTTTCAGAGTCAAGGCGGCGGCGGCGGCGGCGGCGG 704
 OY 807 CAAGATCAACCAAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 863
 DB 705 CGAAGATCTTCAACCAAGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 764
 OY 864 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 923
 DB 765 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 824
 OY 924 CCAAGATCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 983
 DB 825 CAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 884
 OY 984 GAGCGTGAACGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1043
 DB 885 CATGACTTCTTCAACCAAGATCAACAGAGAGAGAGAGAGAGAGAGAGAGAG 944
 OY 1044 CGGAGATCAACCGCTTCAACCAAGATCAACAGAGAGAGAGAGAGAGAGAGAG 1103
 DB 945 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
 OY 1104 CCAAGATCAACCGCTTCAACCAAGATCAACAGAGAGAGAGAGAGAGAGAGAG 1163
 DB 1005 CAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055
 OY 1164 GACCAAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1223
 DB 1056 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1115
 OY 1224 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1283
 DB 1116 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1175
 OY 1284 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1343
 DB 1176 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1232
 OY 1344 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1400
 DB 1233 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1292
 OY 1401 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1460
 DB 1293 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1352
 OY 1461 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1513
 DB 1353 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1412
 OY 1514 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1571
 DB 1413 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1472
 OY 1572 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631
 DB 1473 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1532
 OY 1632 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1691
 DB 1533 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1592
 OY 1692 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1751
 DB 1593 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1652
 OY 1752 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1811
 DB 1653 GAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712

Qy	1812	GNTGATCTGGGGCAAGACCCCGAGTTCCGGCTGCCATCCAGAAAGAGACCTGGGAGAC	1871
Qy	1812	GNTGATCTGGGGCAAGACCCCGAGTTCCGGCTGCCATCCAGAAAGAGACCTGGGAGAC	1871
Db	1713	CAAGAAAGAGATCGAGAGACCCCATCGATGGCGCCATCAGCTGGCTCGACACCAACAGCT	1772
Qy	1872	CTGGTGGACCGACTACTGGCAGGCGACCTGGATCCCGAGTGGAGATTCTGAAACACCC	1931
Db	1773	CGCCGAGGCGGACGAGTTCGAGAGACA--ACATGAAAGAGCTGGAGGGCATCTGCAACCC	1829
Qy	1932	CCCCCTGTGAACTGTGTGTATCAGACTGGAAGAGAGCCCATCATCGGCGCCGAGACTT	1991
Db	1830	GATCATCGCCCAAGATGTATCCAGGGGCCCGGCGCCGACATGGCCGAGGCGATGACGAGGA	1889
Qy	1992	CTACGTGACGGCGCGCCGCAACCGCGAGAACCAAGATCGGCAAGGCCGGCTA	2042
Db	1890	CGCGCCGCGGGCGGCGAGCGGCGCAGAGGCCCCAAAGATCGAAGAGGTTCACACTA	1940

LOCUS	1781 bp	mRNA	linear	EST 03-JUN-2005
CX099623				
DEFINITION	RcM0260 A normalized whole-life-cycle cDNA library of rice <i>Oryza sativa</i> (indica cultivar-group) cDNA clone EI057014, THREBB09, EI057004, EI062D05, EI027114, EI070G16, EI08 5', mRNA sequence.			

ACCESSION	CK093623
VERSION	CK093623.1
KEYWORDS	GI:66912775
SOURCE	EST.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

Bakaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Erihartoideae; Oryzeae; Oryza.

REFERENCE	AUTHORS	TITLE
1 (pages 1 to 1/81)	Zhang, J., Feng, Q., Jin, C., Qiu, D., Zhang, L., Xie, K., Yuan, D., Han, B., Zhang, Q. and Wang, S.	Features of the expressed sequences revealed by a large-scale

JOURNAL
analyses of ESTs from a normalized cDNA library of the elite indica
rice cultivar Minghui 63
Plant J. 42 (5), 772-780 (2005)

PUBMED 15918889
 COMMENT
 Contact: Wang S
 National Key Laboratory of Crop Genetic Improvement
 Huazhong Agricultural University

Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shipingwang@hotmail.com
Seq primer: 17.

FEATURES	Location/Qualifiers
source	1. .1781

```

/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/strain="indica"
/cultivar="Minghui 63"
/db_xref="taxon:33946"
/clcge="E1057014, THREEB09, E1057004, E1062D05, B1027114, E1070G16, E108"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of rice"
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; The library is constructed based on the strategy of saturation hybridization with genomic DNA using rice cultivar Minghui 63. This library consists of cDNA from 15 directionally cloned cDNA libraries constructed with different tissues from 9 developmental stages."

```

	ORIGIN
Query Match	3.3%;
Best Local Similarity	Score 82.6; DB 8;
Matches 449; Conservative	Pred. No. 0.00014;
	Mismatches 584; Indels 6; Gaps 1

QY 231 GGACCTGGACCTTTCCTCCCGAGGGCAGAGGCCCGCGCAGTTTCCCAACGAGCAACACCGCCCA 250

Db 194 GGACCGCGGTGCTCGACGGGTGTCTCGCCCAAGAACCCCGACAGCAAGTGGCTTGGCGAGAC 253

QY 291 CAGCCCTCACAGCCCGAGCTGCGAGTGTGCGCGGCAACACCCCGCAGCGAGGCTCGGCTC 350

Db 254 GTGCACCAAGACCAACATGTGTATGTGTTCGGCGAGATCAACCAAGCCACCGTCCA 313

QY 351 CGAGGGCCAGGGCACCTTGAACTTCCCCCAGATCACTCTGTGGCAGCGCCCTCGTGAG 410

Db 314 CTACGAGAAATCTGTCCGGACAACCTGCGCGGCATCGGCTTGTGTCCAGCAGCTCGG 373

QY 411 CATCAAGGTGGCGGCGCAGATCAAGAGGCGCTGTCTGACACCGCGCGCAGACACCGT 470

Db 374 CCTCGACCGCAGCCGCTGCAAGTGTGTCTCAACATCGAGCAGCAGTGCGCCGACATCGC 433

Db 434 GAGGGGGGTGCA CGGCCCACTTACCAAGGCCCCGAGGAGATCGGGCCCGGCAACGAGG 492
Qy 531 CTTCAATCAAGGTGCGCCAGTACGACCAAGTCTGTATCGAGATTGGGGCAAGAGGCCAT 590
Db 494 CCACATGTTCCGCTAGCCGACCGAGACGACCCCGAGTGTATGCCCTCGACGACGTCCT 553

[illegible]

Db
UY
651
614

CAGGCGCGA
CGGCAAGAC
CCCGAGTCA
CCCGAGTAC
CCTCAACGA
CGCGCGCGC
CATGGT

673

Dy
711 GCCCGCATGGAACGGCCCCAGAGTGAAGATGTGGCCCTGACCGAGAGAAATCAAAGC 770

Db
674 CCCCGTCGGCTCAACAGTCTCTTCTTTCAACCAGACAAGAGACCGTCAACACGA 733

Oy 771 CCTGACCGCCATCTGTGGAGAGATGAGAGAGGGCAAGATCACCAAGATCGGCCCGCA 833

Dh 774 CAGCTCTCCCCC-----GACTCTCABGAGAGTCTTTTCBAAAGGCGCATCTTCCGAACA 780

[illegible]

DB 785 GTACCTCGAGGAGAAACATCTTCACCTCCACCCCTCCGAGCCGATCCGAGC
QY 891 GCTGNGAGCTTCCGAGACTGAAACAAGCACCAGACTTCTGGAGGNGCAGTGG
DB 848 GCCCCAGGCGAGCCGCGCTACCCGCGGAGATCATATGACACTAAGCGGAGATG 907

951 CATCCCCACCCCGCCGCGCTGAAGAGAAGAAGACGTAACCGTCTGACGTTGGCGA 10

Db	908	GGGCGCGCACGGGGGCGGCGCTTCTTCGGCAAGGACCCGACCAAGGTCCAGACCGAAGGG	96
Qy	1011	CGCCTACTTTCAGCGGTGCCCCCTTGACAGGACATTCCGCGMAGTACACGCGCTTTCACCATCCC	107
Db	968	CGCTTACATTCGCGCAGGCGAGCGCCCAAGAGCATCTGTGCGCAGGGGCGCTCCGCCCGCGCTG	102
Qy	1071	CAGCATCAACACGAGAACCCCGCGCATCCGCTTACAGTACACAGTGTGCGCCAGGGGTG	113
Db	1028	CATCTGCGAGGTGTCTTACGCCATCGGCGTCCCGAGCGCGCTCTCCGTGTTCGTGCATTC	108
Qy	1131	GAAGGCGAGCCCCGAGCATTTTTCAGAGCAGCATGACCAAGATCTCTGAGGCCCTTCCGCGC	113
Db	1088	CTACGCGCACCGGCGAAGTCCCCGACAGAGAGATCTCTCAAGATCGTCAAGGAGAACTTCGA	114
Qy	1191	CGCGAACCCCGAGATCGTATCTACAGTACATGAGACGACTGTACGTGGGCGAGGACCT	122
Db	1148	TTTTCAGGCCCGGATGATGACATCTCAACTCTGACCTTAAAGGGGCGGCAACCGGTTTAT	122

QY 1251 GGAGATCGGCAGACCGC 1269
||| ||| |||
Db 1208 CAAGACCGCGCGCTACGCG 1226

RESULT 13
 Cw922203/13
 LOCUS Cw922203 1060 bp DNA linear GSS 24-FEB-2005
 DEFINITION EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA
 Acanthamoeba castellanii genomic clone EDCAR29, genomic survey
 sequence.
 ACCESSION Cw922203.1 GI:60228146
 VERSION Cw922203
 KEYWORDS GSS.
 SOURCE Acanthamoeba castellanii
 ORGANISM Acanthamoeba castellanii
 Buxariota; Acanthamoebidae; Acanthamoeba.
 REFERENCE 1 (bases 1 to 1060)
 Anderson, I.J. and Loftus, B.J.
 TITLE Gene discovery in the Acanthamoeba castellanii genome
 JOURNAL Unpublished (2004)
 COMMENT Contact: Iain Anderson
 The Institute for Genomic Research (TIGR, www.tigr.org)
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-795-7949
 Fax: 301-838-0208
 Class: shotgun.
 FEATURES
 source location/Qualifiers
 1..1060
 /organism="Acanthamoeba castellanii"
 /mol_type="genomic DNA"
 /strain="Neff"
 /db_xref="taxon:5755"
 /clone="EDCAR29"
 /clone_1lb="A. castellanii, 6-8 kb library from total
 genomic DNA"
 /note="Vector: pHD29"

Query Match 3.3%; Score 80.8; DB 10; Length 1060;
 Best Local Similarity 43.9%; Pred. No. 0.00029;
 Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

261 CGAGTTCCCGACGAGCAGACCGCGCCAAAGCCCGACCGCGAGCTGAGGTGG 320
 955 COTGGCGCGGCGACGTCGCCAACAACCTCCAGAGCCCGCGTCATCGCCTGACATGG 896
 331 CGGCGCAACAACCCCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
 895 CGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 836
 381 GATCAACCTGTGGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 440
 835 CAGAGAGGTCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 776
 441 CCTGCTGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
 775 GCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 716
 501 GAAGCCCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
 715 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 656
 561 CCGTATCGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 620
 655 GGAAGAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 596
 621 GAACATCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 680
 595 CGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 536
 681 CCCCATCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 740
 535 GCGCATCATGAGACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
 741 GTGGCCCTCGACCGAGGAGAGATCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 797
 475 CCGCTTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416

QY 798 GAAGAGGCGCAAGATCAACCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 857
 DB 415 GCAAGTCGACCTCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
 QY 858 CATCAAGAGAGAGACAGACCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 917
 DB 355 GGTGAGGCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 296
 QY 918 GCGACCGCAGGACTTCTGGGAGGTGACCTGGGCGCTCCCGACCGCGCGCGCGCG 977
 DB 295 GCGCATCAAGAGAGATCAAGAGAGAGACTTCTGGGCGCTCAAGAGAGAGATCGAGCGCGCTCA 236
 QY 978 GAAGAGAGCGGTGACCGGTGCGACGTGGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1037
 DB 235 GCGCGCGCTACAGAGAGAGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
 QY 1038 GGAATTGCGCAAGTACACCGCGCTTCAACATCCCGAGCATCAACAGAGAGCGCGCGCAT 1097
 DB 175 CGAGGACCTCAAGGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 116
 QY 1098 CCGCTACCAAGTACACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1149
 DB 115 CGATCTGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 64

RESULT 14
 LOCUS C2216254 892 bp DNA linear GSS 10-FEB-2005
 DEFINITION A1AA-eat23g02.b1 Ancylostoma caninum whole genome shotgun library
 (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
 ACCESSION C2216254
 VERSION C2216254.1 GI:59229909
 KEYWORDS GSS.
 SOURCE Ancylostoma caninum (dog hookworm)
 ORGANISM Ancylostoma caninum
 Buxariota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 REFERENCE 1 (bases 1 to 892)
 Mitreva, M., McCarter, J.P., Pape, D., Rieger, R., Tsagarisavlili, R.,
 Ronko, I., Martin, J., Wylie, T., Dante, M., Meyer, R., Messina, D.,
 Waterston, R.H., Clifton, S.W. and Wilson, R.
 TITLE Genome Survey sequences from the parasitic nematode Ancylostoma
 caninum
 JOURNAL Unpublished (2004)
 COMMENT Contact: Mitreva M
 Washington University in St. Louis
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: nematode@wustl.edu
 Genomic DNA provided by John Hawdon (mtmj@wumc.edu) DNA
 sequenced by Washington University Genome Sequencing Center
 Class: shotgun.
 FEATURES
 source location/Qualifiers
 1..892
 /organism="Ancylostoma caninum"
 /mol_type="genomic DNA"
 /strain="Baltimore"
 /db_xref="taxon:29170"
 /dev_stage="Adult"
 /lab_host="G810"
 /clone_1lb="Ancylostoma caninum whole genome shotgun
 library (A1AAGSS 001)"
 /note="Vector: pOTW13; Site 1: Batx1; Site 2: Batx1;
 Ancylostoma caninum genomic DNA was randomly sheared,
 end-repaired and size fractionated to enrich for 2-4 kb
 fragments. Genomic DNA was provided by John Hawdon
 (mtmj@wumc.edu) at George Washington University,
 Sequencing by Washington University Genome Sequencing
 Center, St. Louis, MO."

ORIGIN

Query Match 3.3%; Score 80.6; DB 10; Length 892;
 Best Local Similarity 46.5%; Pred. No. 0.0031;
 Matches 334; Conservative 0; Mismatches 379; Indels 6; Gaps 2;

1656 GATTCGAGAGCAGGCGCAGACAGACAGTGGACCTTACGATCTACCAAGAGCCCTTCAAGAA 1715
 5 GATTCGAGTCTGAGACCTTCAGACGACGTCACACGACGTCACCACTTCAGCGT 64

1716 CCTGAGACCGGCAAGTACGCCAAGATGGCGACCGGCCACACCAACGACGTGAAGAGCT 1775
 65 CGACGACACCAAGCACAACCACTTCACCAACCACTTCGACGACGACGACGACGACGACG 124

1776 GACCGAGCGCTGACAGAGATGCGCATGAGAGCATGTGATCTGAGGGCAGACCCCA 1835
 125 CTAGCAGCTCAAGCAGCAGTCAACGACGCTCAACGACGATGACGTTCGACGACAC 184

1836 GTTCCGCTGCCATTCAGAGAGACCTGGGAGACCTGTGACCGACCTACCTGGAGGC 1895
 185 CAGGACCAACTAGTCCACCACTTCGACGACGACGACGACGACGACGACGACGACGACG 244

1896 CACTGATATCCCGAGTGGAGTTCTGTGAACACCCCGCTGGTGAAGCTGTGTACCA 1955
 245 CAGCAGCAGCTCAACAGACGTCACACGACGCTGACGTGACGACGACGACGACGACG 304

1956 GCTGAGAAAGAGCCCATCATCTGCGCGCCGACACTTCTAGTGAACGCGCGCCCAACCG 2015
 305 CCACTGTCACCACTTCGACGACGACGACGACGACGACGACGACGACGACGACGACG 364

2016 CAGACCAAGATGGGACGCGCTTACGACGACGACGACGACGACGACGACGACGACGACG 2075
 365 CGTCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 424

2076 CCGACCGAGACCAACCA---CCAGAAAGCCGAGCTGAGGCGCATCGCTGAGC 2132
 425 CAGACCGAGACTACTAGTCAAGCAGCTCAAGCAGCAGCTCAACGACGACGACGACG 484

2133 GAGCAGCGGACGAGGTGAATCTGTACCGACGACGACGACGACGACGACGACGACGACG 2192
 485 GCACTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 544

2193 GCGCCAGCCGACCAAGAGCGAGCGAGCTGTGAACCAAGTATGAGAGCGTGTACCA 2252
 545 CGACTCTAGCTGAGAGCGTCAACGACGACGACGACGACGACGACGACGACGACGACG 604

2253 GAGGAGAGAGGTACTGAGCTGGGTCGCGGCCCAAGAGGATCGGCGCAACGACGA 2312
 605 CGTCGACGACCAAGCACAACGCTTCACCAACGACGACGACGACGACGACGACGACG 664

2313 GATCGACAGCTGTGACAGAGGCGCATCCGAAAGTGTCTTCTGAGCGGCAATCGATG 2371
 665 GCTCG---AGCAGGTCAACGACGACGTCACGACGACGACGACGACGACGACGACG 720

RESULT 15
 BM320864 1132 bp mRNA linear EST 03-JAN-2002
 LOCUS rockefeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
 DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
 sequence.
 ACCESSION BM320864 GI:18055270
 VERSION BM320864.1
 KEYWORDS EST.
 SOURCE Mastigamoeba balamuthi
 ORGANISM Mastigamoeba balamuthi
 Eukaryote; Pelobiontida; Mastigamoebidae; Mastigamoeba.
 REFERENCE 1 (bases 1 to 1132)
 Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sengen, C. W.,
 Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., and
 Philippe, H.
 The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

PUBMED 11830664
 COMMENT Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockefeller.edu
 Insert Length: 1132 Std Error: 0.00
 POLY-A-No. Location/Qualifiers

FEATURES
 source
 1. .1132
 /organism="Mastigamoeba balamuthi"
 /mol_type="mRNA"
 /strain="ATCC 30984"
 /db_xref="taxon:108607"
 /clone_11b="Mastigamoeba balamuthi lambda ZAP II Library"
 /note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.2%; Score 80.2; DB 3; Length 1132;
 Best Local Similarity 45.3%; Pred. No. 0.0037;
 Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

139 CCGCCCCCGCAGAGAGGCTGTGAGAGTCCGCGCAAGAGGCGCCACAGATGAAGAC 198
 79 CCGCGGAGGAGGAGAGCGATACCGCGCGCGCCGACGATGATCCAGAGCAAGACAG 138

199 TGCACGAGCGCCAGGCACTTCTTCCGCGAGACCTGACCTTCCCGAGGCAAGGCC 258
 139 TCAACAGCCCAAGATACGCTTCGTC-----GTCCGCTTCAACAAAGGAGCATCGTC 192

259 CCGGATTCCTCCAGAGAGAGACCGCGCCACAGCCCAACAGCCCGAGCTGACAGTG 318
 193 TGCAGATGCTTACCGCAAGATGACGCGGACCAATCTGCGCGCGCTTACTGCGAC 252

319 CCGGCGACCAACCCCGCAGAGGCGCGCGCCGACGCGCGCAGGAGCACTTGAATCCCC 378
 253 GAGCTACCGCTTGGCGCTGACGCTGCGCTGACCACTTACCGCGCGCGCTTACGAGGT 312

379 CAGATCACCTGTGGAGGCGCGCGCTGTGAGATCAAGGTGGCGCGCAATCAAGAG 438
 313 GCGCTGCTGCGCGCGCGCGCTGTGAGAGGCTCACTTCACTCAAGTACGAGGT 372

439 GCGCTGCTGACACCGCGCGCGCGCACACCTGCTGAGAGAGATGAGCTGCGCGCAAG 498
 373 GTCAAGAGGTCAACCGCGAGAGCTTCAACGTCGAGAGAGCTTCAACGAGCGCGCGCT 432

499 TGGAGCCCAAGATGATGCGCGCGCATCGGCGCTTCAATCAAGGTGCGCGCATGACAG 558
 433 TTCAGGCGCTGCTGACGCTGCGCTGTGTCGACCTTCACTGCGCGCGCGCTTGGCC 432

559 ATCTGATGAGATCTGCGCGCAAGAGGCGCATCGGACGCTGTGATCGGCGCGCACCC 618
 493 GCGCTCAAGGAGATGTCGACGCGCGCGCTCAAGCTCC---CAGAGGAGCCGCGCTTC 549

619 GTGAACATCATCGCGCGCAACATGTCGACCGGCTGCGACCTTGAATCCCATC 678
 550 GTTCGCTTCAACCGCGCAAGAGAGCTTCAACGCGCGCTTCTCGCAAGTACATTT 609

679 AGCCCATGAGAGCGGTGCGCGCGTGAAGCTGAAGCCCGGAGATGAGCGGCGCGCAAG 728
 610 GCGGCGCAAGTGGCGCGCGCTGATGAGCTTCTTCAAGAGAGAGAGAGCGCGCGCTTGC 669

739 CAGTGGCCCTTGAACGAGAGAGATCAAGGCGCTTGAACGCGCATCTGCGAGAGATGAG 798
 670 CGCAGATTCGCGGCTTACGCAAGAGAGGAGTGTCAACGCGCGAGATGCTGAGAGATCTAC 729

799 AAGAGAGGCAAGTATCAACAGATGCGCGCGCGAGAACCCCTTCAACACCCCGTGTGCC 858
 730 ACCGAGGCCCAAGAGAGATCCGCGCGAGCTTCAAGCTTGTCTTCCCAAGCGCGCTGAG 789

859 ATCAAGAGAGAGAGAGCAACCAAGTGGCGCAAGCTGTGAGATCTTCCGCGAGCTGAACAG 918
 790 CCGAGGCGCGCAAGCCCAAGCAGCTGGGCGAGCGAGCTGACGATACCAAGCGCAG 849

Tue Jan 10 16:35:49 2006

us-09-610-313b-30.rst

Page 14

QY 919 CGCACCAGGACTTCTGGAGGTGACGTGGGCATCCCCACCCCGCC 966
DB 850 AACCGCGTGGCCAGAGAAGTCCGCGTGGGCTAACCCGACGCCCC 897

Search completed: December 31, 2005, 02:31:20
Job time : 8727.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 11955.3 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463
Sequence: 1 ggcagcagccaccatcgccga.....gggctagcccggtgaatcc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 segs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb_ha: 2: gb_in: 3: gb_env: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pr: 9: gb_ro: 10: gb_sts: 11: gb_sy: 12: gb_un: 13: gb_vi: 14: gb_hlg: 15: gb_pl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2463	100.0	2463	6	AX455915 Sequence
2	2442.2	99.2	2463	6	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916 Sequence
4	2046	83.1	2306	6	BD263704 Improved
5	2046	83.1	2306	6	CQ870574 Sequence
6	2046	83.1	2306	6	AR373387 Sequence
7	2027.6	82.3	9166	6	AX427930 Sequence
8	2025.2	82.2	2312	6	BD263706 Improved
9	2025.2	82.2	2312	6	CQ870576 Sequence
10	2025.2	82.2	2312	6	AR373389 Sequence
11	2019.2	82.0	2300	6	BD263705 Improved
12	2019.2	82.0	2300	6	CQ870575 Sequence
13	2019.2	82.0	2300	6	AR373388 Sequence
14	2019.2	82.0	9788	6	AX427936 Sequence
15	2005.2	81.4	9169	6	AX427931 Sequence
16	2001.8	81.3	9194	6	AX427926 Sequence
17	2000.2	81.2	9194	6	AX427925 Sequence
18	2000.2	81.2	12411	6	AX427927 Sequence

19	1996.8	81.1	9785	6	AX427938 Sequence
20	1993.4	80.9	9167	6	AX427933 Sequence
21	1993.4	80.9	9170	6	AX427928 Sequence
22	1993.4	80.9	9782	6	AX427935 Sequence
23	1993.4	80.9	9783	6	AX427934 Sequence
24	1993.4	80.9	9790	6	CS078943 Sequence
25	1993.4	80.9	9792	6	AX427932 Sequence
26	1993.4	80.9	36066	6	CS078956 Sequence
27	1991.8	80.9	9189	6	AX427921 Sequence
28	1988.8	80.7	4353	11	AF287353 Synthetic
29	1972.2	80.1	3015	6	AX455946 Sequence
30	1955.6	79.4	3009	6	AX455987 Sequence
31	1942.6	78.9	4319	6	BD263639 Improved
32	1942.6	78.9	4319	6	CQ870498 Sequence
33	1942.6	78.9	4319	6	AR373322 Sequence
34	1932	78.4	4352	11	AF287352 Synthetic
35	1932	78.4	4352	11	AY789509 Synthetic
36	1886.6	76.6	2577	6	AX457088 Sequence
37	1879	76.3	4343	6	AX306428 Sequence
38	1878.8	76.3	2305	6	BD263702 Improved
39	1878.8	76.3	2305	6	CQ870572 Sequence
40	1878.8	76.3	2305	6	AR373385 Sequence
41	1875	76.1	4341	6	AX306429 Sequence
42	1865.4	75.7	1977	6	AX455954 Sequence
43	1857.8	75.4	1965	6	AX455952 Sequence
44	1857.2	75.4	7344	6	CS078948 Sequence
45	1852.4	75.2	7897	6	AX427923 Sequence

ALIGNMENTS

RESULT 1	AX455915	2463 bp	DNA	linear	PAT 06-JUL-2002
LOCUS	AX455915	Sequence 31 from Patent WO0204493.			
DEFINITION	AX455915				
ACCESSION	AX455915	GI:21714900			
VERSION	AX455915.1				
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	1				
AUTHORS	zur Megede, J., Barnett, S. W., Engelbrecht, S. and van Rensburg, B.				
TITLE	Polynucleotides encoding antigenic hiv type c polypeptides.				
JOURNAL	Patent: WO 0204493-A 31 17-JAN-2002;				
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)					
FEATURES					
source	1..2463				
ORIGIN					
Query Match	100.0%; Score 2463; DB 6; Length 2463;				
Best Local Similarity	100.0%; Pred. No. 26-202;				
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	GTGAGGCCCACTGGCCGAGGCGCATGAGCCAGGCCCAAGCGCCCAATCTCGATTCAG	60		
DB	1	GTGAGGCCCACTGGCCGAGGCGCATGAGCCAGGCCCAAGCGCCCAATCTCGATTCAG	60		
QY	61	CGAGCACTTCAAGGGGCCCAAGCGCATCATGAGTCTTCAACTGCGGCAAGAGGGC	120		
DB	61	CGAGCACTTCAAGGGGCCCAAGCGCATCATGAGTCTTCAACTGCGGCAAGAGGGC	120		
QY	121	CACATGCGCCGCACTGCGCGCCCGCCGCAAGAGGCTGTGGAAGTGCGGCAAGAG	180		
DB	121	CACATGCGCCGCACTGCGCGCCCGCCGCAAGAGGCTGTGGAAGTGCGGCAAGAG	180		
QY	181	GGCAGCAATGAAGAGTGCACCGAGCGCCAGGCCCACTTCTTCCGCGAGAGACTGGCC	240		

181 GGCACAGATGAAGACTGCAACGAGCGCCAGGCAACTTCTCGGAGGAGACTGAGC 240
QY 241 TTCCCCCAAGGCAAGGCGCGGAGATTCCCCAGAGAGAGAACCGGCGCAACGCCCCACC 300
Db 241 TTCCCCCAAGGCAAGGCGCGGAGATTCCCCAGAGAGAGAACCGGCGCAACGCCCCACC 300
QY 301 AGCCGAGACTGCAAGGTGCGGCGGAGCAACCCCGGAGAGAGCGGCGCGGAGCGCCAG 360
Db 301 AGCCGAGACTGCAAGGTGCGGCGGAGCAACCCCGGAGAGAGCGGCGCGGAGCGCCAG 360
QY 361 GGCACCTTGAACCTTCCCCAGATCAACCTGTGAGCAGCGCCCTGTGAGCATCAAGGTG 420
Db 361 GGCACCTTGAACCTTCCCCAGATCAACCTGTGAGCAGCGCCCTGTGTGAGCATCAAGGTG 420
QY 421 GCGCGGCAATCAAGAGAGCGCTGTGAGCAACCGGCGGAGAGCAACCGTGTGAGAGAG 480
Db 421 GCGCGGCAATCAAGAGAGCGCTGTGAGCAACCGGCGGAGAGCAACCGTGTGAGAGAG 480
QY 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGCGGCGCATGCGGCGCTTCATCAAG 540
Db 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGCGGCGCATGCGGCGCTTCATCAAG 540
QY 541 GTGCGGCAATCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATGCGGCAACGTG 600
Db 541 GTGCGGCAATCAAGATCTGTGATGAGATCTGCGGCAAGAGGCGCATGCGGCAACGTG 600
QY 601 CTGATGCGGCGGCAACCCCGGAGCAATCATGCGGCGGAGCAAGTGTGAGCGCTGAGC 660
Db 601 CTGATGCGGCGGCAACCCCGGAGCAATCATGCGGCGGAGCAAGTGTGAGCGCTGAGC 660
QY 661 ACCCTGAATCTTCCCATCAGCCCATCAGAGACCGTGCCTGTGAGGTGAGCGCGGAGT 720
Db 661 ACCCTGAATCTTCCCATCAGCCCATCAGAGACCGTGCCTGTGAGGTGAGCGCGGAGT 720
QY 721 GACGCGCCCAAGGTGAAGAGTGAAGCCCTGAACGAGAGAGATCAAGGCGCTGACCGGC 780
Db 721 GACGCGCCCAAGGTGAAGAGTGAAGCCCTGAACGAGAGAGATCAAGGCGCTGACCGGC 780
QY 781 ATCTGAGAGAGATGAGAGAGAGGAGAGATCAACAGATGCGGCGCGGAGAGCCCTTAC 840
Db 781 ATCTGAGAGAGATGAGAGAGAGGAGAGATCAACAGATGCGGCGCGGAGAGCCCTTAC 840
QY 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCCGAGACTGAACAGGCGCAACCAAGACTTCTGAGAGGTGAGAGTGTGAGAGTCC 960
Db 901 TTCCGAGACTGAACAGGCGCAACCAAGACTTCTGAGAGGTGAGAGTGTGAGAGTCC 960
QY 961 CCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 AGCGTGCCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 AGCGTGCCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 CCGAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 CCGAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 GAGATGATATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GCGGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 GCGGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320

1261 GCGGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1501 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1561 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1681 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 AGCATGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1801 AGCATGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 ACCCGCGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 ACCCGCGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 ACCCTTCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 1981 ACCCTTCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 AGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400

QY 2401 TACGTGGGAGGCGGCGCTTAGATTCATTAAGCTTCCGGGCTAGACCGGTGA 2460
DB 2401 TACGTGGGAGGCGGCGCTTAGATTCATTAAGCTTCCGGGCTAGACCGGTGA 2460
QY 2461 TTC 2463
DB 2461 TTC 2463
RESULT 2
AX455914 2469 bp DNA linear PAT 06-JUL-2002
LOCUS Sequence 30 from Patent WO0204493.
DEFINITION AX455914
ACCESSION AX455914.1 GI:21714899
VERSION
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source 1. 2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"
ORIGIN
Query Match 99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 1.2e-200;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGAGCCGACCAATGCGCCGAGCCATAGCGCCAGCCGCAACATCTCTGATGAG 60
DB 1 GTGAGCCGACCAATGCGCCGAGCCATAGCGCCAGCCGCAACATCTCTGATGAG 60
QY 61 CGGAGCACTTCAAGGGGCCCCCAAGCGCATCATCAATGCTTCACTGCGGCAAGAGGGC 120
DB 61 CGGAGCACTTCAAGGGGCCCCCAAGCGCATCATCAATGCTTCACTGCGGCAAGAGGGC 120
QY 121 CACATGCGCCGCACTGCGCGCGCCCGGCAAGAGGCTGTGAAGTGGCGGCAAGAG 180
DB 121 CACATGCGCCGCACTGCGCGCGCCCGGCAAGAGGCTGTGAAGTGGCGGCAAGAG 180
QY 181 GGCACACAGATGAAGAGCTGCAACGAGCGCCAGGCCAATCTTCCGCGAGGATGAGC 240
DB 181 GGCACACAGATGAAGAGCTGCAACGAGCGCCAGGCCAATCTTCCGCGAGGATGAGC 240
QY 241 TTCCCCCAAGGAGCGCGCGAGTTTCCCAAGGAGCAAGCCGCGCAACAGCCCCACC 300
DB 241 TTCCCCCAAGGAGCGCGCGAGTTTCCCAAGGAGCAAGCCGCGCAACAGCCCCACC 300
QY 301 AGCGGAGAGCTGAGGTGCGCGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGAG 360
DB 301 AGCGGAGAGCTGAGGTGCGCGCGCAACCCCGGAGCGAGCGCGCGCGCGCGCGAG 360
QY 361 GGCACCTGAACTTCCCGCAGATCACTCTGTGCAAGCGCCCTGTGTGAGCATCAAGTG 420
DB 361 GGCACCTGAACTTCCCGCAGATCACTCTGTGCAAGCGCCCTGTGTGAGCATCAAGTG 420
QY 421 GCGGCGCAAGATCAAGAGGCGCTGTGAGCAACGCGCGCGCAACAGCCGTGTGAGAG 480
DB 421 GCGGCGCAAGATCAAGAGGCGCTGTGAGCAACGCGCGCGCAACAGCCGTGTGAGAG 480
QY 481 ATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATGGGCGGCTTCATCAAG 540
DB 481 ATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATGGGCGGCTTCATCAAG 540

QY 541 GTGGCCAGTACAGACAGATCTGATTCGAGTCTGCGCAAGAGGCCATGGGCAACCGTG 600
DB 541 GTGGCCAGTACAGACAGATCTGATTCGAGTCTGCGCAAGAGGCCATGGGCAACCGTG 600
QY 601 CTGATCGGCCCCCAAGCGCGTGAACATCATCGCGCCCAACATGCTGACCCAGCTGGCTGC 660
DB 601 CTGATCGGCCCCCAAGCGCGTGAACATCATCGCGCCCAACATGCTGACCCAGCTGGCTGC 660
QY 661 ACCCTGAATCTTCCCATAGCGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATG 720
DB 661 ACCCTGAATCTTCCCATAGCGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATG 720
QY 721 GACGCGCCCCAAGAGTGAAGAGTGGCCCCCTGACACCGAGAGAAATATCAAGGCCCTTGAACCGCC 780
DB 721 GACGCGCCCCAAGAGTGAAGAGTGGCCCCCTGACACCGAGAGAAATATCAAGGCCCTTGAACCGCC 780
QY 781 ATCTGAGAGAGATGAAGAGAGGCAAGATCAACCAAGTGGGCCCGGAGAACCCCTTAC 840
DB 781 ATCTGAGAGAGATGAAGAGAGGCAAGATCAACCAAGTGGGCCCGGAGAACCCCTTAC 840
QY 841 AACACCCCGTGTTCGCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AACACCCCGTGTTCGCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGTGAGGTGAGTCACTGGGCTATCCCCCACC 960
DB 901 TTCCCGAGCTGAACAAGCGCACCCAGGACTTCTGTGAGGTGAGTCACTGGGCTATCCCCCACC 960
QY 961 CCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 CCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 AGCGTCCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AGCGTCCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AACGAGACCCCGGCAATCCGCTACAGATCAACAGTGTCTGCGCGGCTGAAGAGGAGAGC 1140
DB 1081 AACGAGACCCCGGCAATCCGCTACAGATCAACAGTGTCTGCGCGGCTGAAGAGGAGAGC 1140
QY 1141 CCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 CCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGATCTACCA-----GGCCCCCTGTAGTGTGGGAGCGACCTGAGATGAGC 1254
DB 1201 GAGATGATCTACCA-----GGCCCCCTGTAGTGTGGGAGCGACCTGAGATGAGC 1254
QY 1255 CAGCACCGGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGCTGGGGCTTCAACACC 1314
DB 1255 CAGCACCGGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGCTGGGGCTTCAACACC 1314
QY 1315 CCGGACAAAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
DB 1315 CCGGACAAAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
QY 1321 CCGGACAAAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 CCGGACAAAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1375 GACAAAGTGAACGCTGAGCGCATCGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
DB 1375 GACAAAGTGAACGCTGAGCGCATCGAGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
QY 1435 ATTCAGAACTGTGTGGGCAAGCTGAATCTGGGCGCAGCTGAATCTTACCCCGCATCAAGTG 1494
DB 1435 ATTCAGAACTGTGTGGGCAAGCTGAATCTGGGCGCAGCTGAATCTTACCCCGCATCAAGTG 1494
QY 1441 ATTCAGAACTGTGTGGGCAAGCTGAATCTGGGCGCAGCTGAATCTTACCCCGCATCAAGTG 1500
DB 1441 ATTCAGAACTGTGTGGGCAAGCTGAATCTGGGCGCAGCTGAATCTTACCCCGCATCAAGTG 1500
QY 1495 CCGGAGCTGTGCAAGTGTGTGCGCGCGCAAGAGCGCTGAACGCAATCTGAGCCCTGAGCC 1554
DB 1495 CCGGAGCTGTGCAAGTGTGTGCGCGCGCAAGAGCGCTGAACGCAATCTGAGCCCTGAGCC 1554
QY 1501 CCGGAGCTGTGCAAGTGTGTGCGCGCGCAAGAGCGCTGAACGCAATCTGAGCCCTGAGCC 1560
DB 1501 CCGGAGCTGTGCAAGTGTGTGCGCGCGCAAGAGCGCTGAACGCAATCTGAGCCCTGAGCC 1560
QY 1555 GAGAGAGCGAGCTGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
DB 1555 GAGAGAGCGAGCTGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
QY 1561 GAGAGAGCGAGCTGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GAGAGAGCGAGCTGAGCTGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620

QY 1615 GTGTACTAGACCCGAGCAAGGACCTGTGCGGAGATCCAGAACGAGGCCACGACAG 1674
DB 1621 GTGTACTAGACCCGAGCAAGGACCTGTGCGGAGATCCAGAACGAGGCCACGACAG 1680
QY 1675 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACCCCAAG 1734
DB 1681 TGGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACCCCAAG 1740
QY 1735 ATGGCAGCCGCGCAACCAAGAGCTGAAAGAGAGCTGACCGGAGCGGTGAGAAAGATCCGC 1794
DB 1741 ATGGCAGCCGCGCAACCAAGAGCTGAAAGAGAGCTGACCGGAGCGGTGAGAAAGATCCGC 1800
QY 1795 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGAG 1854
DB 1801 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGAG 1860
QY 1855 ACCTGGAGAGCTGTGTGACCGCATCTGTGGAGGCACTGTGATCCCGAGTGGAGATTTC 1914
DB 1861 ACCTGGAGAGCTGTGTGACCGCATCTGTGGAGGCACTGTGATCCCGAGTGGAGATTTC 1920
QY 1915 GTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGTGAGAAAGAGCCCATTCATCCGC 1974
DB 1921 GTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGTGAGAAAGAGCCCATTCATCCGC 1980
QY 1975 GCGGAGACCTTCTAGGTGACCGGCGCCGCAACCGGAGACCAAGATCCGCAAGGCGGAC 2034
DB 1981 GCGGAGACCTTCTAGGTGACCGGCGCCGCAACCGGAGACCAAGATCCGCAAGGCGGAC 2040
QY 2035 TAGGTGACCGACCGGCGCGGCAAGAGATGTGAGCTGTGACCGGAGACCAACCAAGAG 2094
DB 2041 TAGGTGACCGACCGGCGCGGCAAGAGATGTGAGCTGTGACCGGAGACCAACCAAGAG 2100
QY 2095 ACCGAGCTGCAAGGACATCCAGCTGTGCGCTGTGAGAGCAAGGCGAGAGGTGAACTTCG 2154
DB 2101 ACCGAGCTGCAAGGACATCCAGCTGTGCGCTGTGAGAGCAAGGCGAGAGGTGAACTTCG 2160
QY 2155 ACCGACAGCAAGTACCGCTGTGGCATCATTCAGAGCCGACCGCAAGAGAGAGAGAG 2214
DB 2161 ACCGACAGCAAGTACCGCTGTGGCATCATTCAGAGCCGACCGCAAGAGAGAGAGAG 2220
QY 2215 CTGTGTGACAGATCATTCAGAGAGCTGATCAAGAAAGAGAGAGGTGACTGTGAGTGGTG 2274
DB 2221 CTGTGTGACAGATCATTCAGAGAGCTGATCAAGAAAGAGAGAGGTGACTGTGAGTGGTG 2280
QY 2275 CCGGCGCAACAGGAGTGGCGCGGCAAGAGAGATGCAAGAGTGTGAGAGAGAGAG 2334
DB 2281 CCGGCGCAACAGGAGTGGCGCGGCAAGAGAGATGCAAGAGTGTGAGAGAGAGAG 2340
QY 2335 CGCAAGTGTCTGTCTGTGACCGGAGATCATTCAGAGAGAGAGTGTGATCTTACAGTACAG 2394
DB 2341 CGCAAGTGTCTGTCTGTGACCGGAGATCATTCAGAGAGAGAGTGTGATCTTACAGTACAG 2400
QY 2395 GACCTGTGAGTGGAGAGCGGCGGCTGTGAGATGATTAAGCTTCCCGGCGCTAGCACC 2454
DB 2401 GACCTGTGAGTGGAGAGCGGCGGCTGTGAGATGATTAAGCTTCCCGGCGCTAGCACC 2460
QY 2455 GGTGAAATTC 2463
DB 2461 GGTGAAATTC 2469

RESULT 3
AX455916 2457 bp DNA linear PAT 06-JUL-2002
LOCUS AX455916
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS
SOURCE
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS 1 zur Megele, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

TITLE polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL polypeptides and uses thereof
PATENT: WO 0204493-A 32 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source
1. 2457
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMMM"
ORIGIN
Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 4e-200;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGAGCGCACCATGTGGCGGAGCCATGAGCCACGACGACCAATCTGTATGCG 60
DB 1 GTGAGCGCACCATGTGGCGGAGCCATGAGCCACGACGACCAATCTGTATGCG 60
QY 61 CGGAGCACTTCAAGAGGCGCCCAAGGAGCATCATAGTCTTCAACTGTGGGAGAGAGGCG 120
DB 61 CGGAGCACTTCAAGAGGCGCCCAAGGAGCATCATAGTCTTCAACTGTGGGAGAGAGGCG 120
QY 121 CACATGCGCCGCACTGCGCGCGCCCGCAAGAGAGGCTGTGAAAGTGCGGCAAGAG 180
DB 121 CACATGCGCCGCACTGCGCGCGCCCGCAAGAGAGGCTGTGAAAGTGCGGCAAGAG 180
QY 181 GCGCACAGATGAAAGAGATGCAACGAGCGCCAGGCCAATCTTTCGCGAGAGACTGTGCG 240
DB 181 GCGCACAGATGAAAGAGATGCAACGAGCGCCAGGCCAATCTTTCGCGAGAGACTGTGCG 240
QY 241 TTCCCGCAAGGAGAGGCGCGGAGTTCCCGAGGAGCAAGAACCGGCGCAAGGCCCAAC 300
DB 241 TTCCCGCAAGGAGAGGCGCGGAGTTCCCGAGGAGCAAGAACCGGCGCAAGGCCCAAC 300
QY 301 AGCGCGAGCTGCAAGAGTGGCGGCGCAACACCGCGAGAGCGGCGCGGCGAGCGCGAG 360
DB 301 AGCGCGAGCTGCAAGAGTGGCGGCGCAACACCGCGAGAGCGGCGCGGCGAGCGCGAG 360
QY 361 GCGACCTGAACTTCCCGCAAGATCAACCTGTGAGAGAGGCGCGCTGTGTGAGATCAAG 420
DB 361 GCGACCTGAACTTCCCGCAAGATCAACCTGTGAGAGAGGCGCGCTGTGTGAGATCAAG 420
QY 421 GCGCGCGAGATCAAGAGAGGCGCTGTGAGACACCGGCGCGAGCAACCTGTGTGAGAGAG 480
DB 421 GCGCGCGAGATCAAGAGAGGCGCTGTGAGACACCGGCGCGAGCAACCTGTGTGAGAGAG 480
QY 481 ATGAGCTGTGCGCGGCAAGTGAAGGCCCAAGATGATCGGCGGAGATCGGCGGCTTCAATCAAG 540
DB 481 ATGAGCTGTGCGCGGCAAGTGAAGGCCCAAGATGATCGGCGGAGATCGGCGGCTTCAATCAAG 540
QY 541 GTGCGCGAGTACAGACAGATCTGTATGAGATCTGCGGCAAGAGAGGCAATCGGCAAC 600
DB 541 GTGCGCGAGTACAGACAGATCTGTATGAGATCTGCGGCAAGAGAGGCAATCGGCAAC 600
QY 601 CTGATCGGCGCCCAACCGCGTGAACATCATTCGCGCGCAACATCTGACCCGAGCTGGCGTGC 660
DB 601 CTGATCGGCGCCCAACCGCGTGAACATCATTCGCGCGCAACATCTGACCCGAGCTGGCGTGC 660
QY 661 ACCCTGAACTTCCCGCATGAGCCCATGAGACCGTGTGAGAGCTGAGAGCCCGGAGATG 720
DB 661 ACCCTGAACTTCCCGCATGAGCCCATGAGACCGTGTGAGAGCTGAGAGCCCGGAGATG 720
QY 721 GACGCGCCCAAGAGTGAAGAGAGTGGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
DB 721 GACGCGCCCAAGAGTGAAGAGAGTGGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
QY 781 ATTCGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATGTGGCGCGGAGAACCTCTAC 840
DB 781 ATTCGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATGTGGCGCGGAGAACCTCTAC 840
QY 841 AACACCCCGGTGTGCGCATCAAGAAAGAGACAGACCAAGTGGCGGAGAGCTGTGTGAGC 900

Db	841	AAACACCCCGCTGTTCGCATTCMAAGAAAGGACGACCAAGTGGCCGCAAGCTGGTGGAC	900
OY	901	TTCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGGGAGGTCAGCTGGGCAATCCCGAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGGGAGGTCAGCTGGGCAATCCCGAC	960
OY	961	CCCGCCGGCTTGAAGAAAGAAAGAGCGGTGACCGGTGCTGACGTGGGCGACCGCTTACTTC	1020
Db	961	CCCGCCGGCTTGAAGAAAGAAAGAGCGGTGACCGGTGCTGACGTGGGCGACCGCTTACTTC	1020
OY	1021	AGCGTGGCCCTTGGACGAGGACTTTCGCGAAGTACACCGCTTACCAATCCCGACGATCAC	1080
Db	1021	AGCGTGGCCCTTGGACGAGGACTTTCGCGAAGTACACCGCTTACCAATCCCGACGATCAC	1080
OY	1081	AACGAGACCCCGGCAATCCGCTTACAGTACACGTGGTGGCCCGAGGGCTGGAAAGGAGAC	1140
Db	1081	AACGAGACCCCGGCAATCCGCTTACAGTACACGTGGTGGCCCGAGGGCTGGAAAGGAGAC	1140
OY	1141	CCGAGCACTTTCAGAGAGCAGTACCAAGATCTTGAGCCCTTCCGCGCCCGCAACCC	1200
Db	1141	CCGAGCACTTTCAGAGAGCAGTACCAAGATCTTGAGCCCTTCCGCGCCCGCAACCC	1200
OY	1201	GAGATCTGATCTTACCAAGCCCTCTGTACGTGGGCAAGCCGACTTGGAGATTCGGCTACAC	1260
Db	1201	GAGATCTGATCTTACCAAGCCCTCTGTACGTGGGCAAGCCGACTTGGAGATTCGGCTACAC	1260
OY	1261	CGCGCAAGATCGAGGAGCTGCGCAAGCACTGTCGTGGGCTTTCACCAACCCCGAC	1320
Db	1261	CGCGCAAGATCGAGGAGCTGCGCAAGCACTGTCGTGGGCTTTCACCAACCCCGAC	1320
OY	1321	AAGAAGCACGAAGAGAGCCCGCTTCTCTGTGATGGGCTTACGAGCTGCACCCCGACAG	1380
Db	1321	AAGAAGCACGAAGAGAGCCCGCTTCTCTGTGATGGGCTTACGAGCTGCACCCCGACAG	1380
OY	1381	TGGAACCGTGCAGCCCATTCAGCTGCGCGAAGAGAGAGCTGGAACCGTGAACGACATCCAG	1440
Db	1381	TGGAACCGTGCAGCCCATTCAGCTGCGCGAAGAGAGAGCTGGAACCGTGAACGACATCCAG	1440
OY	1441	AAGCTGGTGGGCAAGCTGGAACGTGGGCGACCGAGATCTAACCCGCGATCAAGGTCGGCGAG	1500
Db	1441	AAGCTGGTGGGCAAGCTGGAACGTGGGCGACCGAGATCTAACCCGCGATCAAGGTCGGCGAG	1500
OY	1495	CTGTGCAAGCTGCTGCGCGCGCCAGAGCCTCTGAACCGACATCTGTGCCCTTGACCGAGGAG	1554
Db	1495	CTGTGCAAGCTGCTGCGCGCGCCAGAGCCTCTGAACCGACATCTGTGCCCTTGACCGAGGAG	1554
OY	1551	GCCGAGCTGAGCTGGCCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGAGCTGTAC	1614
Db	1551	GCCGAGCTGAGCTGGCCGAGAAACCGCGAGATCTCTGCGCGAGCCCGTGCACGAGCTGTAC	1614
OY	1611	TACGACCCCGACAGAGACTTGTGTGGCCGAGATTCAGAAAGCAGGGCCACAGCAATGGAAC	1674
Db	1611	TACGACCCCGACAGAGACTTGTGTGGCCGAGATTCAGAAAGCAGGGCCACAGCAATGGAAC	1674
OY	1671	TACGAGATCTACAGAGACCCCTTCAAGAACTTGAAAGCCGGCAAGTACGCGCAAGTGGCC	1734
Db	1671	TACGAGATCTACAGAGACCCCTTCAAGAACTTGAAAGCCGGCAAGTACGCGCAAGTGGCC	1734
OY	1741	ACCGCCCAACCAAGAGCTGAAGCACTGAACCGAGGCTGTGCAGAGAGATTCGCATGGAG	1800
Db	1741	ACCGCCCAACCAAGAGCTGAAGCACTGAACCGAGGCTGTGCAGAGAGATTCGCATGGAG	1800
OY	1795	AGCATCTGATCTTGGGGGCAAGACCCCAAGTTCCGCTGGCCATTCMAAGAAAGACCTGG	1854
Db	1795	AGCATCTGATCTTGGGGGCAAGACCCCAAGTTCCGCTGGCCATTCMAAGAAAGACCTGG	1854
OY	1861	GAGACTCTGTGAGACGGAATTAAGAGAGCCACTTGATCCCGAGTGGGAATTGCTGAAC	1920
Db	1861	GAGACTCTGTGAGACGGAATTAAGAGAGCCACTTGATCCCGAGTGGGAATTGCTGAAC	1920
OY	1921	ACCCCGCCCTCTGTGAGACTGTGTACCAAGCTGAGAAAGAGGCCATCATCGCGCCGAG	1980
Db	1921	ACCCCGCCCTCTGTGAGACTGTGTACCAAGCTGAGAAAGAGGCCATCATCGCGCCGAG	1980

Db	1915	ACCCCCCCCCCTGGTGAAGCTGTGTGTACCACTGTAGAGAGAGGCCCATCTGCGGCGCCGAG	1974
Qy	1981	ACCTTTCACGTGTGAACGGCGCGCCGACCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG	2040
Db	1975	ACCTTTCACGTGTGAACGGCGCGCCGACCAACCGCGAGACCAAGATCGGCAAGGCGCGCTACGTG	2034
Qy	2041	ACCGACCGGGGCGGGCGAAGATCTGTGAGCTTGACCGAGACCAACCAACCGAAGAACCTGAG	2100
Db	2035	ACCGACCGGGGCGGGCGAAGATCTGTGAGCTTGACCGAGACCAACCAACCGAAGAACCTGAG	2094
Qy	2101	CTGACAGGCGATCCAGCTGGGCGCTGACAGACAGCGGCGACGAGGTGAACATCTGTGACCGAC	2160
Db	2095	CTGACAGGCGATCCAGCTGGGCGCTGACAGACAGCGGCGACGAGGTGAACATCTGTGACCGAC	2154
Qy	2161	AGCCAGTACGCGCCCTGGGCGATCATTCAGGCGCCAGCGCCGACCAAGACGAGCGAGCTGGTG	2220
Db	2155	AGCCAGTACGCGCGCTGGGCGATCATTCAGGCGCCAGCGCCGACCAAGACGAGCGAGCTGGTG	2214
Qy	2221	AACCAAGATCATTCAGACGAGCTGTGATCAAGAGAGAGGTGTACTCTGAGCTGGGTGTCCCGCC	2280
Db	2215	AACCAAGATCATTCAGACGAGCTGTGATCAAGAGAGAGGTGTACTCTGAGCTGGGTGTCCCGCC	2274
Qy	2281	CACAAAGGCGATTCGGCGGCGCAACGAGCAATGACACAGCTGTGTGAGCAAGGGCATCCGACG	2340
Db	2275	CACAAAGGCGATTCGGCGGCGCAACGAGCAATGACACAGCTGTGTGAGCAAGGGCATCCGACG	2334
Qy	2341	GTGCTGTTCCTGACGCGGCATCGATGTGGCGGATCTGTACTTACACAGTACATGTGACGACTG	2400
Db	2335	GTGCTGTTCCTGACGCGGCATCGATGTGGCGGATCTGTACTTACACAGTACATGTGACGACTG	2394
Qy	2401	TACGTGTGGACAGCGGCGGCGCTTAGGATCGATTAAAAAGCTTCCCGGGGCTTAGCACCGGTGAA	2460
Db	2395	TACGTGTGGACAGCGGCGGCGCTTAGGATCGATTAAAAAGCTTCCCGGGGCTTAGCACCGGTGAA	2454
Qy	2461	TTC 2463	
Db	2455	TTC 2457	

LOCUS	BD263704	2306 bp	DNA	linear	PAT 17--JUL-2003
DEFINITION	Improved expression of HIV polypeptides and production of virus-like particles.				
ACCESSION	BD263704				
VERSION	BD263704.1	GI:33073472			
KEYWORDS	JP 2002533124-A/71.				
SOURCE	Synthetic construct				
ORGANISM	Synthetic construct				
REFERENCE	other sequences; artificial sequences.				
AUTHORS	1 (bases 1 to 2306)				
	Barnett, S., Megede, J.Z., Sriwastava, I., Lian, Y., Hartog, K., Liu, H., Greer, C., Selby, M. and Walker, C.				
	Improved expression of HIV polypeptides and production of virus-like particles				
JOURNAL	Patent: JP 2002533124-A 71 08-OCT-2002;				
COMMENT	CHIRON CORP				
	OS Artificial Sequence				
	PN	JP 2002533124-A/71			
	PD	08-OCT-2002			
	PF	30-DEC-1999	JP 2000591193		
	PR	31-DEC-1998	US 60/114495, 01-DEC-1999	US 60/168471	PI
	SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI KIAN HARTOG.				
	PI HONG LUI CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER				
	C12N5/09, A61K37/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC C12N5/10,				
	PC C12N7/00, C12P21/02, C12N5/00, C12N5/00, A61K37/02	CC			
	Description of Artificial Sequence: FS(-). protomod. RTopt. YM FH				
FT	source	1..2306			key
	Location/Qualifiers				
	/organism='Artificial Sequence'.				
FEATURES	Location/Qualifiers				


```
source 1..2306
/mol_type="synthetic construct"
/db_xref="taxon:32630"

Query Match 83.1%; Score 2046; DB 6; Length 2306;
Beet Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

ORIGIN

170 GGGGCAAGAGGGCCCAAGATGAAAGATGCAACGAGCCAGGCCAACTTCTCCGCG 229
    1 GGGGCGCGCAAGAGCAACCAATGAAGATGCACTGAGAGACAGAGGTATTCTTCCGCG 60
QY AGAAGCTGGCTTCCCGCAGGGCAAGGCGCGAGATTCCCGCAGAGCAAGCCCGCA 289
    61 AGAAGCTGGCTTCTGCAAGGAGAGGCGCGAGATTGCAAGAGCAAGCCCGCA 120
QY AGAGCCCAACAGCCGCGAGCTGCAAGTGTGGGCGCGAGAACAAAGCTGAGCGAG 343
    121 AGAGCCCAACCGCGCGAGCTGCAAGTGTGGGCGCGAGAACAAAGCTGAGCGAG 180
QY CCGGCGCGAGCGCCAGGAGCCCTG-----ACATCCCCAGATCACTTGTGGCAGC 397
    181 CCGGCGCGAGCGCCAGGAGCCCTGAGCTTCAACTTCCCGAGATCACTTGTGGCAGC 240
QY GCGCCCTGGTGAAGATCAAGTGGGCGCGAGATCAAGAGGCGCTGTGCAACCGCG 457
    241 GCGCCCTGGTGAAGATCAAGATGGGCGCGAGATCAAGAGGCGCTGTGCAACCGCG 300
QY CCGAGCAACAGCTGTGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATCG 517
    301 CCGAGCAACAGCTGTGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATCG 360
QY GCGGCAATGGCGGCTTCAATCAAGTGGCGCAAGTGAAGCAAGTCTGATGAGATCTG 577
    361 GCGGCAATGGCGGCTTCAATCAAGTGGCGCAAGTGAAGCAAGTCTGATGAGATCTG 420
QY GGAAGAGGCGCATCGGCAACGATGAGTGGGCGCGAGCCCGTGAAGCATCATCGCGCA 637
    421 GGAAGAGGCGCATCGGCAACGATGAGTGGGCGCGAGCCCGTGAAGCATCATCGCGCA 480
QY ACATGTGACCCAGCTGAGCTGCAACCTTGAATTTCCCATCAGCCCATGAGAGCCGTC 697
    481 ACATGTGACCCAGCTGAGCTGCAACCTTGAATTTCCCATCAGCCCATGAGAGCCGTC 540
QY CCGTGAAGCTGAAGCCCGGCAATGAGTGGCGCAAGTGAAGCATGTGGCCCTGACCGAG 757
    541 CCGTGAAGCTGAAGCCCGGCAATGAGTGGCGCAAGTGAAGCATGTGGCCCTGACCGAG 600
QY AGAAGATCAAGGCGCTGAACCGCATCTGAGAGAGTGAAGAGGCGCAAGATCAACA 817
    601 AGAAGATCAAGGCGCTGAACCGCATCTGAGAGAGTGAAGAGGCGCAAGATCAACA 660
QY AGATCGGCGCCCGAGAACCCCTTCAACACCCCGGTGTTCGCATCAAGAGAGAGCAGCA 877
    661 AGATCGGCGCCCGAGAACCCCTTCAACACCCCGGTGTTCGCATCAAGAGAGAGCAGCA 720
QY AATATGGGCGCCGAGAACCCCTTCAACACCCCGGTGTTCGCATCAAGAGAGAGCAGCA 720
    721 AATATGGGCGCCGAGAACCCCTTCAACACCCCGGTGTTCGCATCAAGAGAGAGCAGCA 780
QY CCAAGTGGCGCAAGCTGTGAGCTTCCGAGAGTGAACAGCGCAACCCAGCACTTGTGG 937
    938 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGAGTGAACAGCGCAACCCAGCACTTGTGG 997
QY AGGTGAGCTGGGCAATCCCGCAACCCCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 997
    781 AGGTGAGCTGGGCAATCCCGCAACCCCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 840
QY TGAAGTGGGCGAGCGCTTCAAGAGTGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 1057
    998 TGAAGTGGGCGAGCGCTTCAAGAGTGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 1057
    841 TGAAGTGGGCGAGCGCTTCAAGAGTGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 900
QY CCTTCAACATCCCGAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
    1058 CCTTCAACATCCCGAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
    901 CCTTCAACATCCCGAGCATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
```

```
QY TGGCCCAAGGCGTGAAGGCGAGGCCCAAGATCTTCCAGAGAGAGATGAACCAAGATCTGG 1177
    1118 TGGCCCAAGGCGTGAAGGCGAGGCCCAAGATCTTCCAGAGAGAGATGAACCAAGATCTGG 1177
    961 TGGCCCAAGGCGTGAAGGCGAGGCCCAAGATCTTCCAGAGAGAGATGAACCAAGATCTGG 1020
QY AGCCCTTCCGCGCGCCGCAACCCCGAGATGTGATCTTCAAGAGAGAGAGAGAGAGAGAG 1237
    1178 AGCCCTTCCGCGCGCCGCAACCCCGAGATGTGATCTTCAAGAGAGAGAGAGAGAGAGAG 1237
    1021 AGCCCTTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY GAGACCTTGAAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
    1238 GAGACCTTGAAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
    1081 GAGACCTTGAAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY GCTGGGCGCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
    1298 GCTGGGCGCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
    1141 GCTGGGCGCTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
    1358 GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
    1201 GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
    1418 GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
    1261 GCTTGAAGCTGCAACCCCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY ACCCGGAGATCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
    1478 ACCCGGAGATCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
    1321 ACCCGGAGATCAAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY ACATGTGCGCCCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
    1538 ACATGTGCGCCCTGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
    1381 AGGTATCTCCCTTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY GCGAGCCGCTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
    1598 GCGAGCCGCTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
    1441 AGGAGCCGCTGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY AGCAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
    1658 AGCAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
    1501 AGCAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY CCGGCAAGTGAAGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1777
    1718 CCGGCAAGTGAAGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1777
    1561 CCGGCAAGTGAAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY CCGTGAAGAGATGCGCATGAGAGAGATGATCTTGGGCGCAAGAGAGAGAGAGAGAGAGAG 1837
    1778 CCGTGAAGAGATGCGCATGAGAGAGATGATCTTGGGCGCAAGAGAGAGAGAGAGAGAGAG 1837
    1621 CCGTGAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY TGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1897
    1838 TGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1897
    1681 TGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
    1898 TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
    1741 TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
    1958 TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2017
    1801 TCCCGAGTGAAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077
    2018 AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077
    1861 AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
    2078 AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2137
    1921 AGATGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY GCGAGTGAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2197
    2138 GCGAGTGAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2197
    1981 TGAAGTGAAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
```

Qy	2198	ACAAAGACGGAAGGAGAGCTGGTTAAACCAATCATTCGAGCAGCTGATCAAGAAAGGAAAG	2257
Db	2041	ACAAAGACGGAAGGAGAGCTGGTTAGGCCAGATCATTCGAGCAGCTGATCAAGAAAGGAAAG	2100
Qy	2258	TGTACTGAGCTGGGTTGCCCGCCCAAGAGGCATTCGGCGGCAAGCAGATTCACAAGC	2317
Db	2101	TGTACTGAGCTGGGTTGCCCGCCCAAGAGGCATTCGGCGGCAAGCAGATTCAGCAAGC	2160
Qy	2318	TGGTGAGCGAAGGCGATCCGCAAGGTGCTGTTCTCTGAGACGGCATGATGAGCGGCATTCGTGA	2377
Db	2161	TGGTGAGCGCGCGGATTCGCAAGGTGCTGTTCTCTGAACCGGATGATGAGCGCGCATTCGTGA	2220
Qy	2378	TCCTACAGATACATGAGACGACCTGTATCGTGGCGACGCGCGGCTCTAGATCGATTTAAAGC	2437
Db	2221	TCCTACAGATACATGAGACGACCTGTATCGTGGCGACGCGCGGCTCTAGATTCGATTTAAAGC	2280
Qy	2438	TTCCCGGGGCTAGCACCGGTGAATTC	2463
Db	2281	TTCCCGGGGCTAGCACCGGTGAATTC	2306

RESULT 5	:		
LOCUS	CG870574	2306 bp	DNA
DEFINITION	Sequence 82 from Patent EP1433851.		linear
ACCESSION	CG870574		
VERSION	CG870574.1	GI:5200090	

ORGANISM of genetic construct
synthetic construct
other sequences; artificial sequences

REFERENCE	1
AUTHORS	Barnett, S., Zumragede, J., Srivastava, J.

Liu, H., Greer, C., Selby, M. and Walker, C.

TITLE Improved expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;

CHIRON CORPORATION (US)	location/qualifiers
1. .2306	

```

/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
fst(-).protmod.RToptc.YM"

```

ORIGIN

Query Match	83.1%	Score 2046;	DB 6;	Length 2306;
Best Local Similarly	93.6%;	Pred. No. 1e-166;		
Matches 2159; Conservative	0;	Mismatches 135;	Indels 12;	Gaps 2;

Oy 170 GCGGCAAGAGAGGCCACCAAGATGAAAGACATCACCAGACGCGCAGGCCCACTTTCTTCGCG 229
 Db 1 GCGGCGCGAAGGACACCAAAATGAAAGATTGACTGAAGACAGGCTTAATTTCTTCCGCG 60
 Oy 230 AGGACTTGCGCTTCCCGCAGGCGCAAGCGCCGCGAGTTCCCGACGACGAGAACCGCGCA 289
 Db 61 AGGACTTGCGCTTCTTGCAAGGCGCAAGGCCCGCGAGTTACGACGCGAGCAACCGCGCA 120
 Oy 290 AACGCGCCACCAAGCGCCGAGCTGCAGTGCAGCGCGCG-----ACAACCCCGCAGCGAG 343
 Db 121 AACGCGCCACCGCGCGAGCTGCAGGTGTGGGGCGCGCGAACAACCGCTGAGCGAG 180
 Oy 344 CCGGCGCGGAGCGCCAGGCGCACCTTG-----AATTTCCCGCAATCACCTTGTGGCAGC 397
 Db 181 CCGGCGCGGAGCGCCAGGCGCACCTTGAGCTTCAACTTCCCGCAGATCACCTGTGGCAGC 240
 Oy 398 GCCCGCTGTGAGCATCAAGTGGGCGCGCCAGATCAAGAGGCGCTGTGCGACACCGGCG 457
 Db 241 GCCCGCTGTGACATCAAGATCGGCGCGCACTCAAGAGGCGCTGTGCGACACCGGCG 300
 Oy 458 CCGACGACACCGTGTGAGGAGATGAGCTGCGCGGCAAGTGAAGGCCAAGATGATCG 517

Db	301	CCGACGACACCGTGTCTGAGAGATGAACCTGCCCGGCGAAGTGGAAACCCAGATGATTCG	360
Qy	518	GCGGCATCCGGGGCTTCATCAAGGTGGCGCAGTACGACAGATCTGATCGAGATCTTGGC	577
Db	361	GCGGGATCCGGGGCTTCATCAAGGTGGCGCAGTACGACAGATCTCCGTGGAGATCTGGC	420
Qy	578	GCAAGAAAGCCATCGGCACCGTGTGATTCGGCCCCACCCCGTGAACATCATCGGCCGA	637
Db	421	GCACCAAGGCCATCGGCACCGTGTGATTCGGCCCCACCCCGTGAACATCATCGGCCGA	480
Qy	638	ACATGCTGAACCCAGCTGGGGCTGCACCCCTGAACTTCCCCCATCAGGCCCATCGAGCCGTGC	697
Db	481	ACCTGCTGAACCCAGATCGGCTGCACCCCTGAATCTTCCCCATCAGCCCATCGAGCCGTGC	540
Qy	698	CCGTGAAGCTGAAGCCCGGCATGAGACCGGCCCAAGGTGAAGCACTGGCCCCCTGACCGAGG	757
Db	541	CCGTGAAGCTGAAGCCCGGGGATGAGACCGGCCCAAGGTCAAGCAGTGGCCCCCTGACCGAGG	600
Qy	758	AGAAAGATCAAGGCCCTGTGTGAGATCTGCAACCGGAGATGAGAGAGAGAGGCAAGTTCAGCA	817
Db	601	AGAAAGATCAAGGCCCTGTGTGAGATCTGCAACCGGAGATGAGAGAGAGAGGCAAGTTCAGCA	660
Qy	818	AGATTCGGCCCCGGAACACCCCTACAAACACCCCGGTTCGTCATCAAGAAAGAGGACAGCA	877
Db	661	AGATTCGGCCCCGGAACACCCCTACAAACACCCCGGTTCGTCATCAAGAAAGAGGACAGCA	720
Qy	878	CCAAGTGGCGCAACTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGG	937
Db	721	CCAAGTGGCGCAACTGTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTTGGG	780
Qy	938	AGGTGCAGCTGGGCAATCCCCACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC	997
Db	781	AGGTGCAGCTGGGCAATCCCCACCCCGCGCTGAAGAGAGAGAGCGTGAACCGTGC	840
Qy	998	TGAGACTGTGGGGAAGCGCTTACTTCAGCGTGGCCCCCTGAGACAGGAATTCGCAAGTACACCG	1057
Db	841	TGAGACTGTGGGGAAGCGCTTACTTCAGCGTGGCCCCCTGAGACAGGAATTCGCAAGTACACCG	900
Qy	1058	CTTTCACCATCCCGCAGCATCAACAAGAGAGACCCCGGCATCCGCTACAGTACAAACGTGC	1117
Db	901	CTTTCACCATCCCGCAGCATCAACAAGAGAGACCCCGGCATCCGCTACAGTACAAACGTGC	960
Qy	1118	TGCCCCAGGGCTGGAAGGGCAGGCCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGG	1177
Db	961	TGCCCCAGGGCTGGAAGGGCAGGCCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGG	1020
Qy	1178	AGCCCTTCCGGCCCGCACAACCCCGAGATCTGTATCTTACCAAGCCCCCTCTGTACTGGGCA	1237
Db	1021	AGCCCTTCCGGCAGACAGAACCCCGACATCTGTATCTTACCAAGCCCCCTCTGTACTGGGCA	1080
Qy	1238	GCGACTGTGAAGATGGGGCCAGCACCGCGCCAGATTCGAGGAGCTGCGCAAGCACTGTCTGC	1297
Db	1081	GCGACTGTGAAGATGGGGCCAGCACCGCACCAAGATTCGAGGAGCTGCGCAAGCACTGTCTGC	1140
Qy	1298	GCTGGGGCTTTCACAACCCCGCACAAGAGACCAAGAGAGGCCCCCTTCTCTGTGATGG	1357
Db	1141	GCTGGGGCTTTCACAACCCCGCACAAGAGACCAAGAGAGGCCCCCTTCTCTGTGATGG	1200
Qy	1358	GCTACGAGCTGACATCCCGCACAAGTGGACCGTGCAGCCCATCATCTGCCCCAGAGAGCA	1417
Db	1201	GCTACGAGCTGACATCCCGCACAAGTGGACCGTGCAGCCCATCATCTGCCCCAGAGAGCA	1260
Qy	1418	GCTGGAACGTGAAGACATTCAGAAAGCTGTGTGGCCAGAGCTGAACCTGGGCCAGCGAGATCT	1477
Db	1261	GCTGGAACGTGAAGACATTCAGAAAGCTGTGTGGCCAGAGCTGAACCTGGGCCAGCGAGATCT	1320
Qy	1478	ACCCCGGCAATCAAGTGCAGCGCTGTGTCAAGCTGTGCGCGCGCCAGAGCCCTTGCACCG	1537
Db	1321	ACCCCGGCAATCAAGTGAAGCAGCTGTGTCAAGCTGTGCGCGCGCCAGAGCCCTTGCACCG	1380
Qy	1538	ACATCTGCTCCCTTGAACCGAGAGGCGCGAGCTGGAAGCTGGCCGAGAACCGCGAGATCTTGC	1597
Db	1381	AGTGTATCCCCCTGACCGAGAGGCGCGAGCTGGAAGCTGGCCGAGAACCGCGAGATCTTGA	1440

```
QY 1598 GCGAGCCCGTGCACGCGGCTGTACTACGACCCCGAGAGAGACTGTGGCCGAGATCCAGA 1657
DB 1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCCGAGAGAGACTGTGGCCGAGATCCAGA 1500
QY 1658 AGCAGGCGCCACGACGAGTGAACCTTACGAGATCTTACGAGAGCCCTTCAAGAACTTGAAGA 1717
DB 1501 AGCAGGCGCCAGGCGCGTGAACCTTACGAGATCTTACGAGAGCCCTTCAAGAACTTGAAGA 1560
QY 1718 CCGGACAGTACGCGCAAGATGCGCACCGGCGCACGAGCGATGAAAGCACTGAACCGAGG 1777
DB 1561 CCGGACAGTACGCGCGCATGCGCGCGCGCGCACGAGCGATGAAAGCACTGAACCGAGG 1620
QY 1778 CCGTGCAGAGATCCGCGCATGAGAGCATGTATCTGGGCGCAAGACCCCGAAGTTCCGCG 1837
DB 1621 CCGTGCAGAGATGAGACCGGAGACATGTATCTGGGCGCAAGATCCCGAAGTTCCAGC 1680
QY 1838 TGCCCATCCAGAGAGAGACTTGGAGAGACTGTGAACGACTTACGAGCGCACCTGGA 1897
DB 1681 TGCCCATCCAGAGAGAGACTTGGAGAGACTGTGAAGTATGAGCGAGCGCACCTGGA 1740
QY 1898 TGCCCGAGTGGAGTTCGTGAACACCCCGCGTGTGAAGCTGTGTATCCAGCTGAGAGA 1957
DB 1741 TGCCCGAGTGGAGTTCGTGAACACCCCGCGTGTGAAGCTGTGTATCCAGCTGAGAGA 1800
QY 1958 AGGAGCCCATCATCGGCGCGCGAGACCTTCTACGTGAGACGCGCGCGCACCGAGAGACA 2017
DB 1801 AGGAGCCCATCATCGTGGCGCGCGAGACCTTCTACGTGAGACGCGCGCGCACCGAGAGACA 1860
QY 2018 AGATGCGGAGGCGCGCTACGTGACCGACCGGCGCGCGAGAGATCTGTAGCCTGAGCCG 2077
DB 1861 AGCTGGGAGAGCGCGCTACGTGACCGACCGGCGCGCGAGAGATGTGTAGCATGCGCG 1920
QY 2078 AGACCAACCAACGAGAGACCGAGCTGAGAGGACATCAAGTGGCGCTGTGAGAGACAGCGCA 2137
DB 1921 AGACCAACCAACGAGAGAGACCGAGCTGAGAGGACATCAAGTGGCGCTGTGAGAGAGCGGC 1980
QY 2138 GCGAGGTGAACATCGTGAACGACGACGAGTACGCGCTGTGAGCATCATCCAGCGCCAGCCG 2197
DB 1981 TGGAGGTGAACATCGTGAACGACGACGAGTACGCGCTGTGAGCATCATCCAGCGCCAGCCG 2040
QY 2198 ACAAGAGCGAGAGCGAGCTGTGAACGAGATCATGAGACGCTGATCAAGAGAGAGAGG 2257
DB 2041 ACAAGAGCGAGAGCGAGCTGTGAACGAGATCATGAGACGCTGATCAAGAGAGAGAGG 2100
QY 2258 TGTAACCTGAGCTGGGCGCGCGCGCGCACAGAGGATGCGCGCGCAACGAGAGATCGCAAGC 2317
DB 2101 TGTAACCTGAGCTGGGCGCGCGCGCGCACAGAGGATGCGCGCGCAACGAGAGATCGCAAGC 2160
QY 2318 TGGTGAAGCAAGGCGATCCGCAAGTGTCTTCTGGAACGCGCATCGATGCGGCGCATCGTGA 2377
DB 2161 TGGTGAAGCGCGCGCATCCGCAAGTGTCTTCTGGAACGCGCATCGATGCGGCGCATCGTGA 2220
QY 2378 TCTACAGTACATGAGACGACTGTGACGTGGGCGAGCGCGCGCTAGAGATGATTAAGC 2437
DB 2221 TCTACAGTACATGAGACGACTGTGACGTGGGCGAGCGCGCGCTAGAGATGATTAAGC 2280
QY 2438 TTCCCGGCGGCTAGCACCGGCTGAATTC 2463
DB 2281 TTCCCGGCGGCTAGCACCGGCTGAATTC 2306

RESULT 6
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 82 from patent US 6602705.
ACCESSION AR373387
VERSION AR373387.1 GI:40075490
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2306)
```

```
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
JOURNAL Particles
Patent: US 6602705-A 82 05-Aug-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source
1..2306
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
QY 170 GCGGCAAGAGGCGCCACGAGATGAGACTGACCGGAGCGGCAAGCTTCTTCGCG 229
DB 1 GCGGCGCGAAGAGACCAATGAAAGATGACATGAGAGAGAGCTAATTTCTTCGCG 60
QY 230 AGGACCTGAGCTTCCCGCAGGCGCAAGGCGCGGAGTTCCCGCAGAGAGAACCGCGCA 289
DB 61 AGGACCTGAGCTTCCCGCAGGCGCAAGGCGCGGAGTTCAAGAGCGAGAGACCGCGCA 120
QY 290 ACAAGCCCAACAGCCGCAAGCTGCAAGTGCAGCGCG-----ACAAGCCGCAAGCGAG 343
DB 121 ACAAGCCCAACAGCCGCAAGCTGCAAGTGCAGCGCGAGAGAGAGAGAGAGAGAGAG 180
QY 344 CCGGCGCGAGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 397
DB 181 CCGGCGCGAGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 398 GCGCGCTGTGAGCATCAAGGTGCGCGCGCGCGAGATCAAGAGAGCGCTGTGAGCAACGCG 457
DB 241 GCGCGCTGTGAGCATCAAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 458 CCGAGCAACCGTGTGAGAGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 517
DB 301 CCGAGCAACCGTGTGAGAGAGATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATCG 360
QY 518 GCGGATGCGGCGCTTCAAGAGTGGCGCGAGTACGACAGATCTGTATCGAGATCTGCG 577
DB 361 GCGGATGCGGCGCTTCAAGAGTGGCGCGAGTACGACAGATCTGTATCGAGATCTGCG 420
QY 578 GCAAGAGAGCGATCGGCGCGCGCGCTGATCGGCGCGCACCGCGTGAACATCATCGCGCA 637
DB 421 GCAAGAGAGCGATCGGCGCGCGCGCGCTGATCGGCGCGCACCGCGTGAACATCATCGCGCA 480
QY 638 ACATGCTGACCGAGCTGGGCTGCGACCTTGAACCTTCCCATGAGCCCATCGAGACCGTGC 697
DB 481 ACCTGCTGACCGAGATCGGCTGCGACCTTGAACCTTCCCATGAGCCCATCGAGACCGTGC 540
QY 698 CCGTGAACCTGAAGCGCGCGAGATGAGCGCGCGCGCAAGTGAAGAGTGGCGCGTGAACGAG 757
DB 541 CCGTGAACCTGAAGCGCGCGAGATGAGCGCGCGCGCGCAAGTGAAGAGTGGCGCGTGAACGAG 600
QY 758 AGAAGATCAAGGCGCTGACCGCGCATCTGCGAGAGATGAGAGAGAGAGAGAGATCAACA 817
DB 601 AGAAGATCAAGGCGCTGAGAGATCTGCAACGAGATGAGAGAGAGAGAGATCAAGA 660
QY 818 AGATGCGCGCGCGAGAACCTTACAAACACCGCGTGTTCGCATCAAGAGAGAGAGAGCA 877
DB 661 AGATGCGCGCGCGAGAACCTTACAAACACCGCGTGTTCGCATCAAGAGAGAGAGAGCA 720
QY 878 CCAAGTGGCGAGAGCTGTGTGAGCTTCCCGAGAGTGAACAGAGCGACCCAGAGCTTCTGG 937
DB 721 CCAAGTGGCGAGAGCTGTGTGAGCTTCCCGAGAGTGAACAGAGCGACCCAGAGCTTCTGG 780
QY 938 AGGTGAGCTGGGCGATCCCGCACCGCGCGCGCTGAAGAGAGAGAGAGAGAGCTGACCGTGC 997
DB 781 AGGTGAGCTGGGCGATCCCGCACCGCGCGCGCTGAAGAGAGAGAGAGAGAGCTGACCGTGC 840
QY 998 TGGAGTGGCGAGCGCTTACCTTCAAGCTGCGCGTGAACAGAGACTTCCGCAAGTACACCG 1057
```

Db 841 TGGACGTGGGCGACGCGTACTTCAAGCGTCCCTGGACAAGACCTTCCGCACTACACCG 900
QY 1058 CCTTCACATCCCGACGATCAACACAGAACCCCGGCAATCGCTACAGATCAACGTGC 1117
Db 901 CCTTACCATCCCGACGATCAACACAGAACCCCGGCAATCGCTACAGATCAACGTGC 960
QY 1118 TGGCCCAAGGGCTGGAAGGGCGGCCGACATCTTCCAGAGCAGATGACCAAGATCTCTG 1177
Db 961 TGGCCCAAGGGCTGGAAGGGCGGCCGACATCTTCCAGAGCAGATGACCAAGATCTCTG 1020
QY 1178 AGCCCTTCGCGCGCCGCAACCCCGAGTGTGATCTACAGAGCGCCCTGTGATGTGAGCA 1237
Db 1021 AGCCCTTCGCGAGCAGAACCCCGACATGTGTATCTACAGAGCGCCCTGTGATGTGAGCA 1080
QY 1238 GCGACCTTGAGATCGGCGACGACCGCGCCAGATCGAGAGCTGCGACAGCCTGCTG 1297
Db 1081 GCGACCTTGAGATCGGCGACGACCGCGCCAGATCGAGAGCTGCGCGACGACCTGCTG 1140
QY 1298 GCTGGGGCTTCAACACCCCGCAACAAGACCAAGAGAGCCCTTCTGTGTATG 1357
Db 1141 GCTGGGGCTTCAACACCCCGCAACAAGACCAAGAGAGCCCTTCTGTGTATG 1200
QY 1358 GCTACAGAGTGAACCCCGCAAGTGAACGCTGAGCCCATGAGCTGCCGAGAGAG 1417
Db 1201 GCTACAGAGTGAACCCCGCAAGTGAACGCTGAGCCCATGAGCTGCCGAGAGAGCA 1260
QY 1418 GCTGACCGTGAACGACATCCAGAGCTGTGTGGCAAGCTGAACCTGGCGCAGATCT 1477
Db 1261 GCTGACCGTGAACGACATCCAGAGCTGTGTGGCAAGCTGAACCTGGCGCAGATCT 1320
QY 1478 ACCCGGCGATCAAGGTGTGGCGACGCTGTGCAAGCTGTGTGGCGCGCCAGGCTTGA 1537
Db 1321 ACCCGGCGATCAAGGTGTGGCGACGCTGTGCAAGCTGTGTGGCGCGCCAGGCTTGA 1380
QY 1538 ACATCGTGGCCCGTGAACCGAGAGGCGAGCTGAGCTGGCGAGAACCGGAGATCTG 1597
Db 1381 AGGTGATCCCTTGAACCGAGAGGCGAGCTGAGCTGGCGAGAACCGGAGATCTG 1440
QY 1598 GCGAGCCCGTGAACCGGCGTGTACTAGACCCCGACAGAGCCTGTGTGGCGAGATCCAG 1657
Db 1441 AGAGGCGCGTGAACCGGCGTGTACTAGACCCCGACAGAGCCTGTGTGGCGAGATCCAG 1500
QY 1658 AGCAGGCGCAACGACGATGAGCCTTACAGAGGCTTCAAGACCTTGAAGA 1717
Db 1501 AGCAGGCGCAACGACGATGAGCCTTACAGAGGCTTCAAGACCTTGAAGA 1560
QY 1718 CCGGCAAGTACCGGCAAGGCGGACCGGCCCAACCGACGCTGAAGCAGCTGAACCGAG 1777
Db 1561 CCGGCAAGTACCGGCGGCGGCGGCGCCACACCAACGCTGAAGCAGCTGAACCGAG 1620
QY 1778 CCGTGAAGATCGCATGAGAGCATGCTGATCTGTGGGCAAGACCCCGAAGTCCGCG 1837
Db 1621 CCGTGAAGATGAGAGCATGCTGATCTGTGGGCAAGATCCCGAAGTCCGAGC 1680
QY 1838 TGGCCATCCAGAGAGAGCCTGTGAGACCTGTGTGAACCTGTGAGAGGCGCACTTGA 1897
Db 1681 TGGCCATCCAGAGAGAGCCTGTGAGAGGCGCTGTGTGAAGTACTGTGAGAGGCGCA 1740
QY 1898 TGGCCATGAGAGGCTGTGTGAACACCCCGCGTGTGTGAAGCTGTGTGAACCTGTGA 1957
Db 1741 TGGCCATGAGAGGCTGTGTGAACACCCCGCGTGTGTGAAGCTGTGTGAACCTGTGA 1800
QY 1958 AGGAGCCCATCATGCGCGCGGAGCCTTCTACGTGAGCGCGCGGCAACCGCGAGACCA 2017
Db 1801 AGGAGCCCATCATGCGCGCGGAGCCTTCTACGTGAGCGCGCGGCAACCGCGAGACCA 1860
QY 2018 AGATCGGCAAGGCGGCTACGTGAGCAGACCGGCGCGGCGGCAAGATGTGTAGCTTGA 2077
Db 1861 AGTGGGCAAGGCGGCTACGTGAGCAGACCGGCGCGGCGGCAAGATGTGTAGCTTGA 1920
QY 2078 AGACCAACCAACGAGAGACCGAGCTGAGAGCAGTCCAGCTGCGCTGTGAGAGACGAGCA 2137
Db 1921 AGACCAACCAACGAGAGACCGAGCTGAGAGCAGTCCAGCTGCGCTGTGAGAGACGAGCA 1980

QY 2138 GCGAGGTGAACATCTGTGACCGACAGCCGATAGCCCTGGGCAATCATCCAGGCCCAAGCCG 2197
Db 1981 TGGAGGTGAACATCTGTGACCGACAGCCGATAGCCCTGGGCAATCATCCAGGCCCAAGCCG 2040
QY 2198 ACAAGAGCGAGGCGGAGCTGTGTGAACAGATCATGAGAGGCTGTCAAGAGAGAGAG 2257
Db 2041 ACAAGAGCGAGGCGGAGCTGTGTGAACAGATCATGAGAGGCTGTCAAGAGAGAGAG 2100
QY 2258 TGTACTGAGTGTGTGCGCGCCCAACAGAGGATCGGCGCAACAGCAGATGACAAAC 2317
Db 2101 TGTACTGAGTGTGTGCGCGCCCAACAGAGGATCGGCGCAACAGCAGATGACAAAC 2160
QY 2318 TGTGTGAGCAGGCGATCGGCAAGGTGTCTTCTGTGAGCGGATCGATGCGCGCATGTGA 2377
Db 2161 TGTGTGAGCAGGCGATCGGCAAGGTGTCTTCTGTGAGCGGATCGATGCGCGCATGTGA 2220
QY 2378 TCTACAGTACATGTGACGACCTGTACGTGTGGGAGAGCGCGCCCTAGATCATTAAGC 2437
Db 2221 TCTACAGTACATGTGACGACCTGTACGTGTGGGAGAGCGCGCCCTAGATCATTAAGC 2280
QY 2438 TTCCCGGGCTAGCACCGGTGAAATTC 2463
Db 2281 TTCCCGGGCTAGCACCGGTGAAATTC 2306

RESULT 7
AX427930 9166 bp DNA linear PAT 20-JUN-2002
LOCUS AX427930
DEFINITION Sequence 168 from Patent W00232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.

REFERENCE
1 Huang, Y. and Nabel, G.J.
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 168 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
Location/Qualifiers
FEATURES
source
1. .9166
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN
Query Match 82.3%; Score 2027.6; DB 6; Length 9166;
Best Local Similarity 91.7%; Pred. No. 2.4e-165;
Matches 2194; Conservative 0; Mismatches 179; Indels 19; Gaps 4;

QY 14 TGGCCAGGCGCATGAGCGGCAACGAGGCGCAACATCTGTATGACGCGGCAACTTCA 73
Db 2961 TGGCCAGGCGCATGAGCGGCAACGAGGCGCAACATCTGTATGAGCGGCGCAACTTCA 3020
QY 74 AGGCGCCCAAGGCGCATCATAGTGTCTTCAACTGTGGGCAAGAGGCGCATGCGCCGCA 133
Db 3021 AGGCGCCCAAGGCGCATCATAGTGTCTTCAACTGTGGGCAAGAGGCGCATGCGCCGCA 3080
QY 134 ACTGCGCGCGCCCGCAAGAGAGGCTGTGAAAGTGTGGGCAAGAGAGGCGCACAGATGA 193
Db 3081 ACTGCGCGCGCCCGCAAGAGAGGCTGTGAAAGTGTGGGCAAGAGAGGCGCACAGATGA 3140
QY 194 AGAAGTGAACCGAGGCGGCGGCAACTTCTTCCGCGAGAGACTGTGCTTCCCGAGGCA 253
Db 3141 AGAAGTGAACCGAGGCGGCGGCAACTTCTTCCGCGAGAGACTGTGCTTCCCGAGGCA 3199
QY 254 AGGCGCGGAGTTCCTCCAGAGCAGAACCGGCGCAACAGCCCGCAAGCGCGGAGCTGC 313
Db 3200 AGGCGCGGAGTTCCTCCAGAGCAGAACCGGCGCAACAGCCCGCAAGAGAGAGCTTC 3259

Db 1561 TGAGACCGGCAAGTACCGCCCGCATGCGGCGGCCCAACACGACGTGAAGACCTGA 1620
QY 1772 CGAGGCGCTGTCAGAGATCGCCATGAGAGCATGTCATCTGCGGCAAGACCCCAAGT 1831
Db 1621 CGAGGCGCTGTCAGAGATCGCCATGAGAGCATGTCATCTGCGGCAAGACCCCAAGT 1680
QY 1832 TCCGCTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
Db 1681 TCAGCTGCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1892 CCTGATATCCCGAGTGGAGAGTCTGTGAACAACCCCTGCTGTGAAGCTGTGTACAGC 1951
Db 1741 CTTGATATCCCGAGTGGAGAGTCTGTGAACAACCCCTGCTGTGAAGCTGTGTACAGC 1800
QY 1952 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
Db 1801 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 2012 AGACCAAGATCGGCAAGGCGGCTACGTGAACCGACCGGCGCGGCAAGAGATCGTAGCC 2071
Db 1861 AGACCAAGATCGGCAAGGCGGCTACGTGAACCGACCGGCGCGGCAAGAGATCGTAGCC 1920
QY 2072 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131
Db 1921 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 2132 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2191
Db 1981 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2192 AGCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2251
Db 2041 AGCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2252 AGAAGGTGATCTGAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2311
Db 2101 AGAAGGTGATCTGAGCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2160
QY 2312 ACAAGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2371
Db 2161 ACAAGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2372 TCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2431
Db 2221 TCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2432 AAAAGCTTCCCGGCGCTAGACACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCCGGCGCTAGACACCGGTGAATTC 2312

RESULT 9
LOCUS CQ870576 2312 bp DNA linear PAT 13-SBP-2004
DEFINITION Sequence 84 from Patent EP1433851.
ACCESSION CQ870576
VERSION CQ870576.1 GI:52000092
KEYWORDS
SOURCE
ORGANISM
1
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
Barnett, S., Zumege, J., Srivastava, I., Lian, Y., Hartog, K.,
Liu, H., Greer, C., Selby, M. and Walker, C.
Improved expression of HIV polyproteins and production of
virus-like particles
Patent: EP 1433851-A 84 30-JUN-2004;
CHIRON CORPORATION (US)
Location/Qualifiers
1..2312
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"

/note="Description of Artificial Sequence:
FS (-).promod.Rlopt (+)"

Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 6.1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 229
Db 1 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 230 AGAAGCTGAGCTTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
Db 61 AGAAGCTGAGCTTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 290 AGAAGCTGAGCTTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 343
Db 121 AGAAGCTGAGCTTCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 344 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 397
Db 181 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 398 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 457
Db 241 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 458 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
Db 301 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 518 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
Db 361 GCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 578 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
Db 421 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 638 AGATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 697
Db 481 AGATGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTG 540
QY 698 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
Db 541 CCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 758 AGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
Db 601 AGAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 818 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 877
Db 661 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 878 CCAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 937
Db 721 CCAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 938 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 997
Db 781 AGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840
QY 998 TGAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
Db 841 TGAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 1058 CTTTACCAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
Db 901 CTTTACCAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 1118 TGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177

```

Db      961  TCCCCAGGAGCTGGAAGGAGCCCGCCATCTTCCAGAGCAGATGACCAAGATCTGG 1020
Qy      1178  AGCCCTTCGCGCCCGCAACCCCGAGATCGTATCTTACA-----GAGCCCTCTGTAGC 1231
Db      1021  AGCCCTTCGCGAGAGAACCCCGCATCGTATCTTACCAAGTACATGACGACCTGTAGC 1080
Qy      1232  TGGGAGCGCACTGAGATTCGCGCAGACCCGCGCCAAATTCGAGAGCTGCGCAAGACC 1291
Db      1081  TGGGAGCGCACTGAGATTCGCGCAGACCCGCGCCAAAGTTCGAGAGCTGCGCAAGACC 1140
Qy      1292  TCGTGGCTGGGGCTTCAACACCCCGCAAGAGACCAAGAGAGCCCGCTTCTGT 1351
Db      1141  TCGTGGCTGGGGCTTCAACACCCCGCAAGAGACCAAGAGAGCCCGCTTCTGT 1200
Qy      1352  GGATGGGCTACAGAGCTGCAACCCCGCAAGTGAACCGTGAAGCCATCGAGCTGCCGAGA 1411
Db      1201  GGATGGGCTACAGAGCTGCAACCCCGCAAGTGAACCGTGAAGCCATCGAGCTGCCGAGA 1260
Qy      1412  AGGAGAGCTGGAACCTGGAACGACATTCAGAAAGCTGGTGGCAAGCTGAACTGGGCCAGCC 1471
Db      1261  AGGAGAGCTGGAACCTGGAACGACATTCAGAAAGCTGGTGGCAAGCTGAACTGGGCCAGCC 1320
Qy      1472  AGATCAACCCCGCAACAGTGGCGCAGCTGTGCAAGCTGGCGGGCGGCCAAGGCC 1531
Db      1321  AGATCAACCCCGCAACAGTGGCGCAGCTGTGCAAGCTGGCGGGCGGCCAAGGCC 1380
Qy      1532  TGAACCACTGTCGTCCTGACCCGAGAGAGCGCCAGAGTGAAGCTGGCGGAGAACCGCGAGA 1591
Db      1381  TGAACCACTGTCGTCCTGACCCGAGAGAGCGCCAGAGTGAAGCTGGCGGAGAACCGCGAGA 1440
Qy      1592  TCGTGGCGAGCCCTGTGCAACGCGGTGTACTACGACCCCAAGAGACCTGTGGCGGAGA 1651
Db      1441  TCGTGAAGAGCGCCGTGCAACGAGTGTACTACGACCCCAAGAGACCTGTGGCGGAGA 1500
Qy      1652  TCCAGAGAGAGCGCCGACGACAGTGAACCTACCAAGATCTTACCGAGAGCCCTTCAAGAAC 1711
Db      1501  TCCAGAGAGAGCGCCGACGAGTGAACCTACCAAGATCTTACCGAGAGCCCTTCAAGAAC 1560
Qy      1712  TGAAGACCGGCAAGTACGCAAGATGCGCACCGGCCACCAACGACGTGAAGCAGCTGA 1771
Db      1561  TGAAGACCGGCAAGTACGCGCGCATGCGCGCGCCCAACGACGTGAAGCAGCTGA 1620
Qy      1772  CCGAGGCGGTGCAAGATTCGCGCATGGAAGCATCGTATCTGGGCGAAGACCCCAAGT 1831
Db      1621  CCGAGGCGGTGCAAGATTCGCGCATGGAAGCATCGTATCTGGGCGAAGATCCCAAGT 1680
Qy      1832  TCGGCTGCGCATTCGAGAGAGACCTGGGAGACCTGTGTGAACCGACTACTGGCAGGCCA 1891
Db      1681  TCAAGCTGCGCATTCGAGAGAGACCTGGGAGACCTGTGTGAAGTACTGGCAGGCCA 1740
Qy      1892  CCTGATTCGCCGAGTGGAGTTGTGTAAACACCCCGCCCTGTGAAAGCTGTGTATCAGC 1951
Db      1741  CCTGATTCGCCGAGTGGAGTTGTGTAAACACCCCGCCCTGTGAAAGCTGTGTATCAGC 1800
Qy      1952  TGAAGAGAGAGCCCATCATTCGCGCGCGAGACCTTCTTACGTGAGACGCGCGCCACCGCG 2011
Db      1801  TGAAGAGAGAGCCCATCATTCGCGCGCGAGACCTTCTTACGTGAGACGCGCGCCACCGCG 1860
Qy      2012  AGACCAAGATTCGGCAAGAGCGGCTTACGTGACCGACCGGGGCGCGCAAGAAAGTGTGACA 1920
Db      1861  AGACCAAGATTCGGCAAGAGCGGCTTACGTGACCGACCGGGGCGCGCAAGAAAGTGTGACA 1920
Qy      2072  TGAACGAGACCAACCAAGAGACCGAGCTGAGAGCCATTCAGTGGCCCTGCAAGAGCA 2131
Db      1921  TGAACGAGACCAACCAAGAGACCGAGCTGAGAGCCATTCAGTGGCCCTGCAAGAGCA 1980
Qy      2132  GCGGCAAGAGTGAACATCTGTGACCGACGCGCATTCGCGCGCATTCAGGCGCC 2191
Db      1981  GCGGCAAGAGTGAACATCTGTGACCGACGCGCATTCGCGCGCATTCAGGCGCC 2040
Qy      2192  AGCCCAACAGAGCAGAGGAGCTGTGAAACCAAGATCATCGAGCAGCTGTATCAAGAGG 2251

```

```

Db      2041  AGCCCGACAGAGCGAGAGCAGAGCTGTGTAGGCCAAGTATCATCGAGCAGCTGATCAAGAGG 2100
Qy      2252  AGAAGTGTACTTGAAGCTGGGTGCGCCGCCCAAGAGGATTCGGCGGCAACGACAGATGG 2311
Db      2101  AGAAGTGTACTTGAAGCTGGGTGCGCCGCCCAAGAGGATTCGGCGGCAACGACAGATGG 2160
Qy      2312  ACAAGCTGTGAGCAAGGGCATCCGCAAGTGTCTGTTCTTGAAGCGGATCATGTGGCGCA 2371
Db      2161  ACAAGCTGTGAGCGCGGCATCCGCAAGTGTCTGTTCTTGAAGCGGATCATGTGGCGCA 2220
Qy      2372  TCGATATCTACCAAGTATCATGAGACGACCTGTACGTGTGGCAGCGCGCCCTAGAGTGAAT 2431
Db      2221  TCGATATCTACCAAGTATCATGAGACGACCTGTACGTGTGGCAGCGCGCCCTAGAGTGAAT 2280
Qy      2432  AAAAGCTTCGCGGGCTAGACACCGGTGAATTC 2463
Db      2281  AAAAGCTTCGCGGGCTAGACACCGGTGAATTC 2312

RESULT 10
AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Mesede,J., Greer,C. and Selby,M.
TITLES Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source
1..2312
/organism="Unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 6,1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

Qy      170  GCGGCAAGAGGCGCAACAGATGAAGACTGCAACGAGCGCCAGCCAACTTCTTCGCG 229
Db      1  GCGGCGCGCAAGAGCAACAAATGAAGATTGCACTGAAGACAGGCTATTCTTCGCG 60
Qy      230  AGGACTGTGCTTTCGCCAGAGGCAAGGCGCGAGTTCCCGACGAGCAAGAACCGGCCA 289
Db      61  AGGACTGTGCTTTCGTGCAAGGCGAAGCGCGAGATTGACAGAGCAAGACCGGCCA 120
Qy      290  ACAAGCCCAACAGCCCGAGCTGCAAGTGCAGCGCG-----ACAACCCCGGAGGAGG 343
Db      121  ACAAGCCCAACAGCCCGAGCTGCAAGTGCAGCGCG-----ACAACCCCGGAGGAGG 180
Qy      344  CCGGCGCGAGCGCGAGGCAACCTG-----AACTTCCCGCAGATCAACCTGTGCGAGC 397
Db      181  CCGGCGCGAGCGCGAGGCAACCTGAGCTTCAACTTCCCGCAGATCAACCTGTGCGAGC 240
Qy      398  GCCCTGTGTAGCATCAAGTGGCGCGCGCATGCAAGAGGCGCTGTGGAACAACCGCGC 457
Db      241  GCCCTGTGTAGCATCAAGTGGCGCGCGCATGCAAGAGGCGCTGTGGAACAACCGCGC 300
Qy      458  CGAGACACCGGCTGTGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATG 517
Db      301  CGAGACACCGGCTGTGAGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATG 360
Qy      518  GCGGCAATCGCGGCTTCAATCAAGTGCAGCATGACCAAGATCTGATGAGATCTGCG 577
Db      361  GCGGCAATCGCGGCTTCAATCAAGTGCAGCATGACCAAGATCTGATGAGATCTGCG 420

```

QY 578 GCAGAGAGGCATCGGCAACCTGTGATCGGCCCCCAACCCCTGTGAACATCATCGGCGCA 637
 DB 421 GCCACAGAGGCATCGGCAACCTGTGATCGGCCCCCAACCCCTGTGAACATCATCGGCGCA 480
 QY 638 ACATGCTGACCCAGGCTGGCTGCACTTGAACCTTCCCATCGAGCCCATCGAGACCGTGC 697
 DB 481 ACCTGCTGACCCAGGCTGGCTGCACTTGAACCTTCCCATCGAGCCCATCGAGACCGTGC 540
 QY 698 CCGTGAAGCTGAAGCCCGGCAATGAGACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGG 757
 DB 541 CCGTGAAGCTGAAGCCCGGCAATGAGACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGG 600
 QY 758 AGAAGATCAAGGACCTTGACCGCATCTGCGAGAGAGTGAAGAGAGGCGCAAGTACCA 817
 DB 601 AGAAGATCAAGGACCTTGAGTGAATCTGACCGAGTGAAGAGAGGCGCAAGTACCA 660
 QY 818 AGATGGGCCCCGAGAACCCCTTACAAACCCCGTGTTCGCTCATGAAGAGAGAGAGCA 877
 DB 661 AGATGGGCCCCGAGAACCCCTTACAAACCCCGTGTTCGCTCATGAAGAGAGAGCA 720
 QY 878 CCAAGTGGCGCAAGGCTGTGGACTTCCGAGAGCTGAAGAGGCGACCCGAGACTTCTGGG 937
 DB 721 CCAAGTGGCGCAAGGCTGTGGACTTCCGAGAGCTGAAGAGGCGACCCGAGACTTCTGGG 780
 QY 938 AGGTGCAAGTGGGCAATCCCCCAACCCCGCGCTGAGAGAGAGAGAGAGGCGTGC 997
 DB 781 AGGTGCAAGTGGGCAATCCCCCAACCCCGCGCTGAGAGAGAGAGAGAGGCGTGC 840
 QY 998 TGGAGGTGGGCGAACCGCTTACGAGTGTGCTTGGAGAGAGAGAGAGAGAGAGAGAGAG 1057
 DB 841 TGGAGGTGGGCGAACCGCTTACGAGTGTGCTTGGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 1058 CCTTCACATCCCGAGCATCAACAGAGACCCCGGCAATCCGCTACAGTCAACAGTGC 1117
 DB 901 CCTTCACATCCCGAGCATCAACAGAGACCCCGGCAATCCGCTACAGTCAACAGTGC 960
 QY 1118 TGGCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1177
 DB 961 TGGCCGAGGCTGGAAGGCGAGCCCGAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1178 AGCCCTTCCGCGCCGCAACCCCGAGATGTGATCTACCA-----GGCCCCCTGTAGG 1231
 DB 1021 AGCCCTTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 QY 1232 TGGGAGAGGCACTGAGAGATCGGCGAGAGACCGGCGCAAGATCGAGAGAGAGAGAGAGAG 1291
 DB 1081 TGGGAGAGGCACTGAGAGATCGGCGAGAGACCGGCGCAAGATCGAGAGAGAGAGAGAG 1140
 QY 1292 TGGCTGCTGGGAGCTTCAACACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
 DB 1141 TGGCTGCTGGGAGCTTCAACACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1352 GGAATGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411
 DB 1201 GGAATGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1412 AGGAGAGGCTGAGACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1471
 DB 1261 AGGAGAGGCTGAGACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1472 AGATCTACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1531
 DB 1321 AGATCTACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1532 TGAACGAGCATCGTGTCTTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1591
 DB 1381 TGAACGAGCATCGTGTCTTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1592 TCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1651
 DB 1441 TCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
 QY 1652 TCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711

DB 1501 TCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 QY 1712 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
 DB 1561 TGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1772 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1831
 DB 1621 CCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 QY 1832 TCCGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1891
 DB 1681 TCAAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 QY 1892 CCGTGAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1951
 DB 1741 CCGTGAATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 QY 1952 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2011
 DB 1801 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 2012 AGACCAAGATCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2071
 DB 1861 AGACCAAGATCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 QY 2072 TGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131
 DB 1921 TGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
 QY 2132 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2191
 DB 1981 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
 QY 2192 AGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2251
 DB 2041 AGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 QY 2252 AGAAGGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2311
 DB 2101 AGAAGGTGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2312 ACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2371
 DB 2161 ACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 QY 2372 TCGTATCTACCAAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2431
 DB 2221 TCGTATCTACCAAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
 QY 2432 AAAAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2491
 DB 2281 AAAAGCTTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340

RESULT 11

BD263705 2300 bp DNA linear PART 17-UTL-2003
 LOCUS Improved expression of HIV polypeptides and production of
 DEFINITION virus-like particles.
 ACCESSION BD263705
 VERSION BD263705.1 GI:33073473
 KEYWORDS UP 200253124-A/72.
 SOURCE synthetic construct
 ORGANISM other sequences; artificial sequences.
 REFERENCE 1 (bases 1 to 2300)
 AUTHORS Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
 Greer,C., Selby,M. and Walker,C.
 TITLE Improved expression of HIV polypeptides and production of
 JOURNAL virus-like particles
 Patent: JP 200253124-A 72 08-OCT-2002;

COMMENT

CHIRON CORP
 OS Artificial Sequence
 PN JP 200253124-A/72
 PD 08-OCT-2002
 PR 30-DEC-1999 JP 2000591193
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471, PI
 SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIYASTAVA, YING LIAN, PI
 KARTIN HARTOG,
 PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC
 C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
 C12N5/10,
 PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
 Description of Artificial Sequence: PS(-). protomod. Rtopc. YMM FH
 Key Location/Qualifiers
 FT source 1..2300
 FT Location/Qualifiers
 1..2300
 /organism="Artificial Sequence".
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN

Query Match 82.0%; Score 2019.2; DB 6; Length 2300;
 Best Local Similarity 93.2%; Pred. No. 2e-164; Mismatches 138; Indels 18; Gaps 3;
 Matches 2150; Conservative 0;

170 GCGGCAAGAGGCGCCAGATGATGAGACTGCAACGAGCGCAGGCACTTCTTCGCG 229
 1 GCGGCGCGAAGAGCAACCAATGAAAGATTGCACTGAGAGAGGCTTAATTCTTCGCG 60
 230 AGGACCTGCGCTTCCCGCAAGGCGAAGCGCCGCGAGTTCCCGAGAGCAAGAACCGCGCA 289
 61 AGGACCTGCGCTTCTGCAAGGCGAAGCGCCGCGAGTTCAAGAGCGAGCAAGACCGCGCA 120
 290 ACAAGCCCCAGACCGCGGAGCTGCAAGTGGCGGCG-----ACAACCCCCGAGAGG 343
 121 ACAAGCCCCAGACCGCGGAGCTGCAAGTGGCGGCGGAGAACCAAGCTGAGAGG 180
 344 CCGGCGCGAAGGCGCGAGGCGACCTG-----AACTTCCCGAGATCAACCTTGGCAG 397
 181 CCGGCGCGAAGGCGCGAGGCGACCTGAGCTTCACTTCCCGAGATCAACCTTGGCAG 240
 398 GCGCCCTGCTGAGCATCAAGTGGCGCGCCAGATCAAGAGGCGCTGCTGAGCAACCGCG 457
 241 GCGCCCTGCTGAGCATCAAGTGGCGCGCAAGCTCAAGAGGCGCTGCTGAGCAACCGCG 300
 458 CCGAGCAACCTGCTGAGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATG 517
 301 CCGAGCAACCTGCTGAGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATG 360
 518 GCGGCAATCGGCGGCTTCAATCAAGTGGCGCGAGTACGACAGATCTGATGAGATCTG 577
 361 GCGGCAATCGGCGGCTTCAATCAAGTGGCGCGAGTACGACAGATCTGATGAGATCTG 420
 578 GCAAGAGGCGCATCGGCAACCTGCTGATCGGCCCAACCCCGTGAACATCATCTGCGCA 637
 421 GCAAGAGGCGCATCGGCAACCTGCTGATCGGCCCAACCCCGTGAACATCATCTGCGCA 480
 638 ACAATGCTGACCGAGCTGGGCTGCACTCTGAATTTCCCATCAAGCCCATGAGACCTG 697
 481 ACCTGCTGACCGAGATCGGCTGCACTCTGAATTTCCCATCAAGCCCATGAGACCTG 540
 698 CCGTGAAGCTGAAGCCCGGATGAGAGGCGCCCAAGAGTGAAGAGTGGCCCTTGAACGAG 757
 541 CCGTGAAGCTGAAGCCCGGATGAGAGGCGCCCAAGAGTGAAGAGTGGCCCTTGAACGAG 600
 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGCGCAAGATCA 817
 601 AGAAGATCAAGGCGCTGATGAGATCTGCAACGAGATGAGAGAGGCGCAAGATCA 660
 818 AGATCGGCGCGAGAACCCCTTCAACACCCCGTGTTCGCATCAAGAAAGACGCA 877

661 AGATCGGCGCGAGAACCCCTTCAACACCCCGTGTTCGCATCAAGAAAGACGCA 720
 878 CCAAGTGGCGAGGCTGCTGAGCTTCCGAGAGCTGAACAAGGCGACCCAGAGCTTGG 937
 721 CCAAGTGGCGAGGCTGCTGAGCTTCCGAGAGCTGAACAAGGCGACCCAGAGCTTGG 780
 938 AGGTCAGTGGCGAGTCCCGGAGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAG 997
 781 AGGTCAGTGGCGAGTCCCGGAGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAG 840
 998 TGAAGTGGCGAGCGCTTCACTTCAAGCTGCGCCCTGAGACGAGAGCTTCCGAGTAC 1057
 841 TGAAGTGGCGAGCGCTTCACTTCAAGCTGCGCCCTGAGACGAGAGCTTCCGAGTAC 900
 1058 CTTTCAACATCCCGAGCATCAACAGAGACCCCGGCGATCCGCTTCAAGTACAGTGC 1117
 901 CTTTCAACATCCCGAGCATCAACAGAGACCCCGGCGATCCGCTTCAAGTACAGTGC 960
 1118 TGGCCGAGGCGTGAAGAGGCGAGCGCCAGAGATCTTCCAGAGAGAGATCAAGATCTG 1177
 961 TGGCCGAGGCGTGAAGAGGCGAGCGCCAGAGATCTTCCAGAGAGAGATCAAGATCTG 1020
 1178 AGCCCTTCCGCGCGCAACCCCGAGATGATCTTCAAGAGCGCCCTGTAAGTGGGCA 1237
 1021 AGCCCTTCCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 1238 GCGACCTGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
 1081 GCGACCTGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 1298 GCTGGGCGCTTCAACACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
 1141 GCTGGGCGCTTCAACACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
 1358 GCTAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
 1199 ---CGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
 1418 GCTGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
 1255 GCTGAGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1314
 1478 ACCCGGCGATCAAGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
 1315 AGCGCGGATCAAGAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
 1538 ACAATGTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
 1375 AGATGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
 1598 GCGAGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
 1435 AGGAGCCCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
 1658 AGGAGGCGCAACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
 1495 AGGAGGCGCAACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
 1718 CCGGCAAGTACCGCAAGATGGCGAGCGCCCAACAGAGAGAGAGAGAGAGAGAGAG 1777
 1555 CCGGCAAGTACCGCGAGATGGCGAGCGCCCAACAGAGAGAGAGAGAGAGAGAGAG 1614
 1778 CCGTGAAGAGATCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1837
 1615 CCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
 1838 TGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1897
 1675 TGGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
 1898 TGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
 1735 TGGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794


```

Db 601 AGAATCATAGAGCCCTGCTGAGATCTTGACCGAGATGAGAGAGGCGCAAGATTCAGCA
Qy 818 AGATGGCCCGGAGAAACCCCTTACAAACACCCCGTGTGGCCATCAAGAAAGAGACAGA
Db 661 AATATGGCCCCGAGAAACCCCTTACAAACACCCCGTGTGTGCTCATCAAGAAAGAGACAGA
Qy 878 CCAAGTGGCGCAAGCTGTGTGACTTCCGAGACTGAAACAAGCGCACCCAGACTTCTGG
Db 721 CCAAGTGGCGCAAGCTGTGTGACTTCCGAGACTGAAACAAGCGCACCCAGACTTCTGG
Qy 938 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTGAGAAAGAAAGAGAGCTGACCGTGC
Db 781 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTGAGAAAGAAAGAGAGCTGACCGTGC
Qy 998 TGGAGTGGGCGACCGCTTACTTCAAGGTGTCCCTTGACAGAGACTTCCGCAAGTACCGG
Db 841 TGGAGTGGGCGACCGCTTACTTCAAGGTGTCCCTTGACAGAGACTTCCGCAAGTACCGG
Qy 1058 CCTTCAACATCCCGAGCATCAACAAAGAGACCCCGGCAATCCGCTTACAGTCAACAGTGC
Db 901 CCTTCAACATCCCGAGCATCAACAAAGAGACCCCGGCAATCCGCTTACAGTCAACAGTGC
Qy 1118 TGGCCCAAGGCTGAGAAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGG
Db 961 TGGCCCAAGGCTGAGAAAGGCGAGCCCGAGCATCTTCCAGAGCAGCATGACCAAGATCTGG
Qy 1178 AGCCCTTCCGCGCCCGCAACCCCGAGATGGTATCTACAGAGCCCGCTTACGTGGGGA
Db 1021 AGCCCTTCCGCGAGAGAAACCCCGAGCATGTATCTACAGAGCCCGCTTACGTGGGGA
Qy 1238 GCGACTTGAGATCGGCGCAGCACCGCGCCAGATCGAGAGCTGCGCAAGCACCTGCTGC
Db 1081 GCGACTTGAGATCGGCGCGAGCACCGGACCAAGATCGAGAGCTGCGCGCGCACCTGCTGC
Qy 1297 GCTGGGGCTTCAACACCCCGCAAGAGAGCACCAAGAGAGCCCGCTTCTGTGTATGG
Db 1140 GCTGGGGCTTCAACACCCCGCAAGAGAGCACCAAGAGAGCCCGCTTCTGTGTATGG
Qy 1357 GCTGGGGCTTCAACACCCCGCAAGAGAGCACCAAGAGAGCCCGCTTCTGTGTATGG
Db 1198 GCTGGGGCTTCAACACCCCGCAAGAGAGCACCAAGAGAGCCCGCTTCTGTGTATGG
Qy 1417 GCTAAGAGCTGACCCCGCAAGTGTGACCTGTGACAGCCCTTCAAGAGCTGCGCGAGAGAGA
Db 1254 GCTAAGAGCTGACCCCGCAAGTGTGACCTGTGACAGCCCTTCAAGAGCTGCGCGAGAGAGA
Qy 1477 GCTGAGCCTGAAAGCATTCAGAAAGCTGTGGGCAAGCTGAACTGTGGGCGAGCGAGATCT
Db 1314 GCTGAGCCTGAAAGCATTCAGAAAGCTGTGGGCAAGCTGAACTGTGGGCGAGCGAGATCT
Qy 1537 ACCCGGAGCATCAAGTGTGCGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCGCTTGACCG
Db 1374 ACCCGGAGCATCAAGTGTGCGCAGCTGTGCAAGCTGTGCGCGCGCAAGGCGCGCTTGACCG
Qy 1597 ACATGTGTGCTGACCTGACCGGAGAGCGGAGCTGAGAGCTGTGCGCGAGAACCCGAGATCTTGC
Db 1434 ACATGTGTGCTGACCTGACCGGAGAGCGGAGCTGAGAGCTGTGCGCGAGAACCCGAGATCTTGC
Qy 1657 GCGAGCCCGTGCACCGCGTGTACTAGACCCCGAGCAAGAGACTGTGGCGCGAGATCCAGA
Db 1494 AGGAGCCCGTGCACCGAGGTGTACTAGACCCCGAGCAAGAGACTGTGGCGCGAGATCCAGA
Qy 1717 AGCAGGCGCACGACAGTGTGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAGAGA
Db 1554 AGCAGGCGCACGAGGCGAGTGTGACCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAGAGA
Qy 1777 CCGGCAAGTACCGCAAGATGCGCACCGCGCACCAACGACGTAAGAGAGCTGTGACCGAGG
Db 1614 CCGGCAAGTACCGCGCATGTGCGCGCGCGCACCAACGACGTAAGAGAGCTGTGACCGAGG
Qy 1837 CCGTGCAGAAAGTGGCATGGCATGGAGCATGTGTACTGTGGGCGAAGACCCCAAGTTCCGCG
Db 1674 CCGTGCAGAAAGTGGAGCACCGAGAGCATGTGTACTGTGGGCGAAGATCCCAAGTTCAACG
Qy 1897 TGGCCATCCAGAAAGAGACCTGTGTGACCGGACTACTGTGCGAGCGCACCTGTGA

```

```

Db 1675 TGGCCATCCAGAAAGAGACTTGGAGAGCGCTGTGTGATGTAGTACTGGAGCGCACCTGGA
Qy 1898 TCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTGAGCTGTGTGATCCAGCTGGAGA
Db 1795 TCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTGAGCTGTGTGATCCAGCTGGAGA
Qy 1958 AGGAGCCCATCATCGCGCGCGAGACCTTCTAGTGTGACGCGCGCGCACCGCGAGAGCA
Db 1795 AGGAGCCCATCATCGTGTGGCGCGCGAGACCTTCTAGTGTGACGCGCGCGCACCGCGAGAGCA
Qy 2018 AGATGGCGAGGCGCGCTTCTGTGACCGACCGGCGCGCGCGAGAAAGATCGTGAAGCTGACCG
Db 1855 AGCTGGGAGAGCGCGCTTCTGTGACCGACCGGCGCGCGCGAGAGTGTGTGACATGCGCG
Qy 2078 AGACCAACAACGAGAAACCGGAGTGTGACCGCATCAAGTGTGCGCTGTGACGAGAGCGCA
Db 1915 ACACCAACAACGAGAAACCGGAGTGTGACCGCATCAAGTGTGCGCTGTGACGAGAGCGCG
Qy 2138 GCGAGTGTGAATCGTGTGACCGACAGCTGAGTGTGAGCTGTGATCCAGGCGCGAGCGCG
Db 1975 TGGAGTGTGAATCGTGTGACCGACAGCTGAGTGTGAGCTGTGATCCAGGCGCGAGCGCG
Qy 2198 ACAAGAGCGAGAGCGAGCTGTGTGAACAGATCATGTGACGAGCTGTGATCAAGAAAGAGAG
Db 2035 ACAAGAGCGAGAGCGAGCTGTGTGAACAGATCATGTGACGAGCTGTGATCAAGAAAGAGAG
Qy 2258 TGTACTGTGAGCTGTGTGAGCGCGCGCAAGAGGATGTGGGCGCAACGAGCATGTGACAGC
Db 2095 TGTACTGTGAGCTGTGTGAGCGCGCGCAAGAGGATGTGGGCGCAACGAGCATGTGACAGC
Qy 2318 TGTGAGAGAGGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
Db 2155 TGTGAGAGGCGGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAG
Qy 2378 TCTACAGTGTACGTGAGCACTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTG
Db 2215 TCTACAGTGTACGTGAGCACTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTGTGAGCTGTG
Qy 2438 TTTCCGCGGCTGTGAGCAACCGGTGAAATTC 2463
Db 2275 TTTCCGCGGCTGTGAGCAACCGGTGAAATTC 2300

RESULT 14
AX427936
LOCUS AX427936 9788 bp DNA linear PAT 20-TUN-2002
DEFINITION Sequence 174 from Patent WO0232943.
ACCESSION AX427936
VERSION AX427936.1 GI:21538023
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang Y. and Nabel G.J.
AUTHORS Modifications of hiv env, gag, and pol enhance immunogenicity for
TITLES Genetic immunization
JOURNAL Patent: WO 0232943-A 174 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1..9788
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Plasmid pVR1012x/s containing HIV genes"

ORIGIN
Query Match 82.0%; Score 2019.2; DB 6; Length 9788;
Best Local Similarity 91.6%; Pred. No. 1,2e-164;
Matches 2190; Conservative 0; Mismatches 178; Indels 24; Gaps 4;

14 TGGCGAGGCGCATGAGCGAGCGCACCGCGCAATCTGTGTGAGCGCGCAACTTGA 73

```

Db	2361	TGCGCGAGGCGCATGAGCAAGTGAACAACCAACCAATCATGATGACGCGACGACACTGCA	3020
Oy	74	AGGGCCCCAAGCGCANTCATCAAGTCTTCAACTGCGCGAAGAGGGCCACATGCGCCGA	133
Db	3021	AGGGCCCCAAGCGCACCATCAAGTGTCTTCAACTGCGCGAAGAGGGCCACTGCGCCGA	3080
Oy	134	ACTGCGCGGCCCCCCGCAAGAGGGCTGTGGAAGTGTGCGCGAAGAGGGCCACATATGA	193
Db	3081	ACTGCGCGGCCCCCCGCAAGAGGGCTGTGGAAGTGTGCGCGAAGAGGGCCACATATGA	3140
Oy	194	AGGACTGCAACGAGCGCGCAGGCGCAACTTCTTCCGCGAGGACCTGTGGCTTCCCCAGGGCA	253
Db	3141	AGGACTGCAACGAGCGCGCAGGCGCAACTTCTTCCGCGAGGACCTGTGGCTTCCCCAGGGCA	3194
Oy	254	AGGCGCGCGAGTTCCCGCAGAGCAAGACCGCGCAACAGCCCGCACCGCGAGCTGC	313
Db	3195	AGGCGCGCGAGTTCTTCTTCAAGAGACACAGGCGCAACAGCCCGCACAGAGAGCTTC	3254
Oy	314	AGGTGCGCGG-----CGACAACTCCCGCAGCGAGGCCGCGCGAGCGCCAGGGCA---	364
Db	3255	AGGTTTGGGGGAAGAGACAACACTCCCTTCAGAAGAGGAGCGCGATGACAAAGAACTG	3314
Oy	365	---CCCTGAATTTCCCCCGACATCAACCTGTGTGGCAGCGCCCTCTGTGTAGCATCAAGTGG	421
Db	3315	TATCTTTAAGCTTCCCTCTAGATCACTTTTGTGGACGACCCCTGTGTCAATATAAGATAG	3374
Oy	422	GGCGCGAGATCAAGAGAGGCCCTGTGTGACACCGCGCGCGACACACCGTGTGAGAGAGA	481
Db	3375	GGGCGCGAGCTAAGAGAGGCCCTTCTTAAGACACCGCGCGCGACACCGTGTGTGAGAGAGA	3434
Oy	482	TGAGCTTGTCCCGGCAAGTGAAGGCCCAAGATGATCGGCGGATCGGCGGCTTCATCAAGG	541
Db	3435	TGAACCTGTCCCGGCGCTGGAAGGCCCAAGATGATCGGCGGCGATCGGCGGCTTCATCAAGG	3494
Oy	542	TGCGCGAGTACAGACCAAGTCTGTATTCAGATCTGTGCGCGAAGAGGCCATGTGGCACCGCTGC	601
Db	3495	TGGCGCAATACAGACCAAGTCTGTATTCAGATCTGTGCGCGCAAGAGGCCATGTGGCACCGCTGC	3554
Oy	602	TGATTCGGGCCCAACCCCGGTGAACATCATTCGCGCGCAACATGCTTACCAGCTGTGGCTGCA	661
Db	3555	TGTGTGGGCCCAACCCCGGTGAACATCATTCGCGCGCAACCTGTCTAACCAGTGTGGCTGCA	3614
Oy	662	CCCTGAATTTCCCATCAAGCCCATTCAGACCGTGTCCGTGAAGCTGAAGCCCGGCAATGG	721
Db	3615	CCCTGAATTTCCCATCAAGCCCATTCAGACCGTGTCCGTGAAGCTGAAGCCCGGCAATGG	3674
Oy	722	ACGGCCCCAAGGTGAAGCATGTGGCCCTTGACCGAGAGAGATCAAGGCTTGAACCGCA	781
Db	3675	ACGGCCCCAAGGTGAAGCATGTGGCCCTTGACCGAGAGAGATCAAGGCTTGTGTGAGA	3734
Oy	782	TCTGTGAGGAGATGAGAGAGAGGCGCAAGATCAACCAAGATGTGGGCCGAGAACCCCTTACA	841
Db	3735	TCTGTGAGGAGATGAGAGAGAGGCGCAAGATCAACCAAGATGTGGGCCGAGAACCCCTTACA	3794
Oy	842	ACACCCCGGTGTGGCATCAAGAGAGAGCAGACCAACCAAGTGTGCGCAAGCTGTGTGACT	901
Db	3795	ACACCCCGGTGTGGCATCAAGAGAGAGCAGACCAACCAAGTGTGCGCAAGCTGTGTGACT	3854
Oy	902	TCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGTGAAGTGTGACGTGGGACTTCCCGCAC	961
Db	3855	TCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGTGAAGTGTGACGTGGGACTTCCCGCAC	3914
Oy	962	CCGCGCGGCTGAAGAGAGAGAGCGAGACCGTGTGAGAGTGTGGGCGAGCGCTTACTTCA	1021
Db	3915	CCGCGCGGCTGAAGAGAGAGAGCGAGACCGTGTGAGAGTGTGGGCGAGCGCGCTTACTTCA	3974
Oy	1022	GGGTGCCCCCTGACAGAGACTTTCGCAAGTACACCGCCTTCAACATCCCGACATCAACA	1081
Db	3975	GGGTGCCCCCTGACAGAGACTTTCGCAAGTACACCGCCTTCAACATCCCGACATCAACA	4034
Oy	1082	ACGAGACCCCGGGCATTCGGCTTACCAATACCAAGTGTGTGCCCCAGGGGCTGAAGGGCAGCC	1141
Db	4035	ACGAGACCCCGGGCATTCGGCTTACCAATACCAAGTGTGTGCCCCAGGGGCTGAAGGGCAGCC	4094

OY	1142	CCAGCATCTTCCAGAGAGATGACCAAGATCTCTGAGGCCCTTCGGGCCGACACCCG	1201
Db	4095	CCGCATCTTCAATGACGATGACCAAGATCTCGAGCCCTTCCGAGACGAAACCCG	4154
OY	1202	AGATCGTATCTCA-----GGCCCCCTGTATCTGAGGCAAGCAGCCTGGAGATCGGC	1255
Db	4155	ACATCTGATCTTACAGTACATGGACACCTGTATCTGAGGAGAGCCTGGAGATCGGCC	4214
OY	1256	AGCACCGGCCAAGATCGAGGAGCTGGCGCAAGACCTGCTGCGCTGGGGCTTCAACACCC	1315
Db	4215	AGCACCGGACCAAGATGAGGAGCTGGCGCAAGACCTGCTGCGCTGGGGCTTCAACACC	4274
OY	1316	CCGACAAAGAACACAGAGAGGCCCCCTTCTGTGAGTGGGCTAGAGCTGACACCCG	1375
Db	4275	CCGACAGAGACACAGAGAGGCCCCCTTCTGTGATGGGCTACAGCTGACACCCG	4334
OY	1376	ACAAATGAGCCGTCGACGCCATCGAGCTGCCTCGAGAGAGAGAGCTTGACCTGTAGACA	1435
Db	4335	ACAAATGGACCGTACAGGCCATCTGTCTGCCTCGAGAGAGACAGCTTGACCTGTAGACA	4394
OY	1436	TTCCAGAACTGTGTGGGCACTGAACTGTGGGCAACCCAGATCTTACCCGGGATCAAGGTGC	1495
Db	4395	TTCCAGAACTGTGTGGGCAAGCTGAACTGTGGGCAACCCAGATCTTACCGGGCATCAAGGTGC	4454
OY	1496	GCCAGCTGTGCAACTGCTGCGCGCGCCCAAGGCCCTTGACCGACATGTCGCCCTTGACCG	1555
Db	4455	GCCAGCTGTGCAACTGCTGCGCGCGCCCAAGGCCCTTGACCGAGATGTCGCCGTGACCG	4514
OY	1556	AGAGAGCCGAGCTGAGAGTGGCCGAGAACCGGCGAGATCTCTGCGGAGCCGTGACCGGC	1615
Db	4515	AGAGAGCCGAGCTGAGAGTGGCCGAGAACCGGCGAGATCTCTGAGAGAGCCCGTGCACGGC	4574
OY	1616	TGTACTACGACCCCAAGCAAGGACCTGTGGGCGAGATCTCAAGACAGGGCCACAGCACGT	1675
Db	4575	TGTACTACGACCCCAAGCAAGGACCTGTATCGCGAGATCTCAAGACAGGGCCAGAT	4634
OY	1676	GGAACCTACAGATCTTACAGGAGGCCCTTCAAGAACTTGAAAGCCGGCAAGTACGCCAAGA	1735
Db	4635	GGAACCTACAGATCTTACAGGAGGCCCTTCAAGAACTTGAAAGCCGGCAAGTACCCCGCA	4694
OY	1736	TGCGCACCGGCCCAACCAACGACTGAAAGGACTGACCGAGGCCGTGCAAGAAATTCGCA	1795
Db	4695	TGAGAGGGGCCCAACCAACGACTGAAAGGACTGACCGAGGCCGTGCAAGAAATTCGCA	4754
OY	1796	TGAGAGCATGTGATCTGGGGCAAGACCCCAAGTTCCGCTCGCTCCATCCAGAGAGAGA	1855
Db	4755	CCGAGAGCATGTGTGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAGAGAGA	4814
OY	1856	CTTGAGAGACTGTGTGAACCGACCTTACTGGCAGGCCAACCCTGATCCCGAGTGGAGTTTCG	1915
Db	4815	CTTGAGAGAGCTGTGTGAACCGAGTACTGGCAGGCCCACTTGATCCCGAGTGGAGTTTCG	4874
OY	1916	TGAACACCCCCCCTGTGTGAAGCTGTGTGATCAAGCTGGAAGAGAGGCCCATCATTCGGCG	1975
Db	4875	TGAACACCCCCCCTGTGTGAAGCTGTGTGATCAAGCTGGAAGAGAGGCCCATCATTCGGCG	4934
OY	1976	CCGAGACCTTCTACGTGAGCGGCGCCGCCAACCGCGAGACCAAGATCGGCAAGGCCGCGCT	2035
Db	4935	CCGAGACCTTCTACGTGAGCGGCGCCGCCAACCGCGAGACCAAGCTGGGCAAGGCCGCGCT	4994
OY	2036	ACGTGATCCGACCCGGGGCCGGCAAGATCTGTGACCTTGAACCGAGACCAACCAAGAGA	2095
Db	4995	ACGTGATCCGACCCGGGGCCGGCAAGAGGGTGTGCTCCCTTGACCGACCAACCAAGAGA	5054
OY	2096	CCGAGCTTCAAGGCATCCAGCTGGCCCTTGACGAGCAGCGCAGCGAGAGGTGAACTGTGA	2155
Db	5055	CCGAGCTTCAAGGCATTCACCTGGCCCTTGACGAGCAGCGGCGCTGAGGTGAACTTCGTGA	5114
OY	2156	CCGACAGCCAGTACGCCCTTGGGCATCATTCAGGCCCAAGCCCGCAACAGACGAGACCGAC	2215
Db	5115	CCGACAGCCAGTACGCCCTTGGGCATCATTCAGGCCCAAGCCCGCAACAGACGAGACCGAC	5174


```
Db 4523 AGAGGCGGAGCTGAGCTGGCCGAGAACCGGAGATCTTGAAGAGCCCGTGCACGCG 4582
Qy 1616 TGTACTAGACCCGACGAGAGACTGTGTGGCCGAGATCCAGAAAGCAGGCGCACGACAGT 1675
Db 4583 TGTACTAGACCCGACGAGAGAGCTGATTCGCCAGATCCAGAAAGCAGGCGCACGAGT 4642
Qy 1676 GGAAGCTACGAGATCTTACGAGAGAGCCCTTCAAGAACTTGAAAGACCGGCAAGTACGCCAGA 1735
Db 4643 GGAAGCTACGAGATCTTACGAGAGAGCCCTTCAAGAACTTGAAAGACCGGCAAGTACGCCCA 4702
Qy 1736 TGCAGACCGGCGGACCAAGCAGAGAGCTGAAGACAGTGAACCGAGGCGGTGCAAGATCGCCA 1795
Db 4703 TGAAGGGCGGCCACACCAAGCAGAGAGCTGAACCGAGGCGGTGCAAGATCGCCA 4762
Qy 1796 TGAAGAGCATCGTGAATCTGAGGCGCAGAGCCCAAGTTCCGCTGCGCCATCCAGAAAGAGA 1855
Db 4763 CGGAGAGCATCTGATCTGAGGCGCAGAGCCCAAGTTCAAGCTGCCATCCAGAAAGAGA 4822
Qy 1856 CCTGGAGAGCCTGTGTGAGCCGACCTACTGAGAGGCGCACCTGATTCGCCAGTGGAGTTGG 1915
Db 4823 CTGGGAGGCGCTGTGTGAGCCGAGTACGAGCACGAGCACCTGATTCGCCAGTGGAGTTGG 4882
Qy 1916 TGAACACCCCCCTGTGTGAAGCTGTGTGATCCAGCTGAGAGAGAGCCCATCATCGGGG 1975
Db 4883 TGAACACCCCCCTGTGTGAAGCTGTGTGATCCAGCTGAGAGAGAGCCCATCATCGGGG 4942
Qy 1976 CCGAGACTTCTTACGTTGAGACGCGGCGCCGACCAACCGAGAGACCAAGATCGGCAAGCGGCT 2035
Db 4943 CCGAGACTTCTTACGTTGAGACGCGGCGCCGACCAACCGAGAGACCAAGATCGGCAAGCGGCT 5002
Qy 2036 ACGTGACCGACCGGCGGCGCGGAGAGAGATCGTGAAGCTTGAACCGAGACCAACCAAGAGAGA 2095
Db 5003 ACGTGACCGACCGGCGGCGCGGAGAGAGTGTGCTGACCGACCAACCAACCAAGAGAGA 5062
Qy 2096 CCGAGCTGACAGGCGCATTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 2155
Db 5063 CCGAGCTGACAGGCGCATTCAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 5122
Qy 2156 CCGAGAGCCAGTACGCGCTGTGAGCATCATCAGGCGCCAGCCAGAGAGAGAGAGAGAGAGAG 2215
Db 5123 CCGAGAGCCAGTACGCGCTGTGAGCATCATCAGGCGCCAGCCAGAGAGAGAGAGAGAGAGAG 5182
Qy 2216 TGGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2275
Db 5183 TGGTGAACAGATCATCGAGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5242
Qy 2276 CCGCCCAACAAAGGAGATCGGCGGCAAGAGAGATCGACAAAGCTGTGAGAGAGAGAGATCC 2335
Db 5243 CCGCCCAACAAAGGAGATCGGCGGCAAGAGAGAGATCGGCTGTGAGAGAGAGAGAGATCC 5302
Qy 2336 GCAAGGTGCTGTTCTGTGAGAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2387
Db 5303 GCAAGGTGCTGTTCTGTGAGAGAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5354
```

Search completed: December 30, 2005, 19:16:26
Job time : 11965.3 secs

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1296.67 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463
Sequence: 1 gtcagccaccacatgcccga.....gggtcagccacggtagattc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2463	100.0	2463	6 ABL39960	ABL39960 Synthetic
2	2463	100.0	2463	12 ADM73765	Adm73765 HIV-1 pol
3	2455.4	99.7	2457	8 ACA03547	Aca03547 Synthetic
4	2455.4	99.7	2457	10 ADCl3265	Adcl3265 DNA of HI
5	2442.2	99.2	2469	6 ABL39959	Ab139959 Synthetic
6	2442.2	99.2	2469	12 ADM73764	Adm73764 HIV-1 pol
7	2436.2	98.9	2457	6 ABL39961	Ab139961 Synthetic
8	2436.2	98.9	2457	12 ADM73766	Adm73766 HIV-1 pol
9	2430.2	98.7	2457	8 ACA03548	Aca03548 Synthetic
10	2430.2	98.7	2457	10 ADCl3266	Adcl3266 DNA of HI
11	2422.6	98.4	2445	8 ACA03546	Aca03546 Synthetic
12	2422.6	98.4	2445	10 ADCl3264	Adcl3264 DNA of HI
13	2415.6	98.1	3930	10 ADCl3230	Adcl3230 DNA of HI
14	2414	98.0	3930	10 ADCl3231	Adcl3231 DNA of HI
15	2414	98.0	3930	10 ADCl3232	Adcl3232 DNA of HI
16	2414	98.0	5184	8 ACA03591	Aca03591 Synthetic
17	2414	98.0	5184	10 ADCl3279	Adcl3279 DNA of HI
18	2383.6	96.8	3531	10 ADCl3234	Adcl3234 DNA of HI
19	2382	96.7	3537	10 ADCl3236	Adcl3236 DNA of HI

20	2381	96.7	5145	8 ACA03521	Aca03521 Synthetic
21	2381	96.7	5145	10 ADCl3233	Adcl3233 DNA of HI
22	2371	96.3	3538	10 ADCl3235	Adcl3235 DNA of HI
23	2322.6	94.3	3607	8 ACA03551	Aca03551 Synthetic
24	2322.6	94.3	3607	10 ADCl3269	Adcl3269 DNA of HI
25	2322.6	94.3	3624	8 ACA03550	Aca03550 Synthetic
26	2322.6	94.3	3624	10 ADCl3268	Adcl3268 DNA of HI
27	2304.4	93.6	3597	8 ACA03549	Aca03549 Synthetic
28	2304.4	93.6	3597	10 ADCl3267	Adcl3267 DNA of HI
29	2142	87.0	2466	8 ACA03542	Aca03542 Synthetic
30	2142	87.0	2466	8 ACCT8506	Acct8506 HIV p2Pol
31	2122.8	86.2	2472	8 ACA03543	Aca03543 Synthetic
32	2122.8	86.2	2472	8 ACCT8507	Acct8507 HIV p2Pol
33	2115.2	85.9	2460	8 ACA03541	Aca03541 Synthetic
34	2115.2	85.9	2460	8 ACCT8505	Acct8505 HIV p2Pol
35	2114.2	85.8	3564	8 ACCT8488	Acct8488 HIV GagPo
36	2114.2	85.8	3564	8 ACCT8489	Acct8489 HIV GagPo
37	2113.6	85.8	4716	8 ACA03522	Aca03522 Synthetic
38	2113.6	85.8	4716	10 ADCl3238	Adcl3238 DNA of HI
39	2109.8	85.7	3999	8 ACCT8484	Acct8484 HIV GagCo
40	2108.2	85.6	3999	8 ACCT8485	Acct8485 HIV GagCo
41	2108.2	85.6	3999	8 ACCT8486	Acct8486 HIV GagCo
42	2108.2	85.6	5283	8 ACA03584	Aca03584 Synthetic
43	2108.2	85.6	5283	8 ACCT8529	Acct8529 HIV TatRe
44	2108	85.6	4713	8 ACA03592	Aca03592 Synthetic
45	2108	85.6	4713	10 ADCl3280	Adcl3280 DNA of HI

ALIGNMENTS

RESULT 1
AB139960
ID ABL39960 standard; DNA; 2463 BP.
XX
XX ABL39960;
XX
XX 15-MAY-2002 (first entry)
XX
XX Synthetic construct PR975YM SEQ ID NO:31.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
XX immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
XX immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
XX
XX Synthetic.
XX
XX WO200204493-A2.
XX
XX 17-JAN-2002.
XX
XX PD
XX 05-JUL-2001; 2001WO-US021241.
XX
XX PF
XX 05-JUL-2000; 2000US-00610313.
XX
XX PR
XX (CHIR) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
XX PI
XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
XX WPI; 2002-154920/20.
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
XX applications including DNA immunization or generation of packaging cell
XX lines, particularly in gene therapy.
XX
XX Claim 1; Fig 9; 23pp; English.
XX
XX The present invention describes expression cassettes comprising a
XX polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
XX type C polypeptides. The expression cassettes comprise any of the HIV
XX type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef

CC (1). (1) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunization, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX

SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 2463; DB 6; Length 2463;
Best Local Similarity 100.0%; Pred. No. 4,4e-297;

Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGACGCGCCACCATGGGCGGAGGCGCATGAGCGGCGACCGAGGCGCAACCTGATGCGAG 60
DB 1 GTGACGCGCGCCACCATGGGCGGAGGCGCATGAGCGGCGACCGAGGCGCAACCTGATGCGAG 60
QY 61 CGCAGCAACTTTCAGAGGCGCCGCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGAGGAGC 120
DB 61 CGCAGCAACTTTCAGAGGCGCCGCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGAGGAGC 120
QY 121 CACATGCGCCGCGCACTGCGGCGCCCGCGCAAGAGGCGTCTGGAAGTGCAGAGAG 180
DB 121 CACATGCGCCGCGCACTGCGGCGCCCGCGCAAGAGGCGTCTGGAAGTGCAGAGAGAG 180
QY 181 GGCACACGATGAGAGGAGTGCACCGAGGCGCGACCACTTCTTCGAGAGGACCTTGCGC 240
DB 181 GGCACACGATGAGAGGAGTGCACCGAGGCGCGACCACTTCTTCGAGAGGACCTTGCGC 240
QY 241 TTCCCGCAGAGGCGAGGCGCGGAGTTCGCCAGCGAGCAGAGACCGGCGCCAGGCCCGCAC 300
DB 241 TTCCCGCAGAGGCGAGGCGCGGAGTTCGCCAGCGAGCAGAGACCGGCGCCAGGCCCGCAC 300
QY 301 AGCGCGGAGCTGCAAGTGCAGGCGCGCAACCCCGCGAGGCGCGCGCGCGCGCGCGCGAG 360
DB 301 AGCGCGGAGCTGCAAGTGCAGGCGCGCAACCCCGCGAGGCGCGCGCGCGCGCGCGCGAG 360
QY 361 GGCACCGCTGAACTTCCCGCGAGTCACTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGAG 420
DB 361 GGCACCGCTGAACTTCCCGCGAGTCACTGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGAG 420
QY 421 GCGCGCGAGATCAAGAGGCGCGTGTGAGCACCGGCGCGCGAGCACCTGTGTGAGAGAG 480
DB 421 GCGCGCGAGATCAAGAGGCGCGTGTGAGCACCGGCGCGCGAGCACCTGTGTGAGAGAG 480
QY 481 ATGAGCGTGCCTGCGGCAAGTGCAGAGCGCGAGATGAGTGCAGGCGCGCGCGCGCGAG 540
DB 481 ATGAGCGTGCCTGCGGCAAGTGCAGAGCGCGAGATGAGTGCAGGCGCGCGCGCGCGAG 540
QY 541 GTCGCGCAGTACGACGAGATCTGATTCGAGATCTGCGGCGAGAGGCGCATCGGCGAGCG 600
DB 541 GTCGCGCAGTACGACGAGATCTGATTCGAGATCTGCGGCGAGAGGCGCATCGGCGAGCG 600
QY 601 CTGATCGGCGCGCGCGCGCGTGAACATCATCGGCGCGCGAGCATCTGAGCCGAGTGGAGTGC 660
DB 601 CTGATCGGCGCGCGCGCGCGTGAACATCATCGGCGCGCGAGCATCTGAGCCGAGTGGAGTGC 660
QY 661 ACCCTGAATCTTCCCATCAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGAGT 720
DB 661 ACCCTGAATCTTCCCATCAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGAGT 720
QY 721 GACGCGCGCGCGAGAGTGCAGGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
DB 721 GACGCGCGCGCGAGAGTGCAGGCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
QY 781 ATCTCGAGAGAGATGAGAGAGAGAGAGAGATCAACAGATGAGCGCGCGCGAGAGAGAGAG 840
DB 781 ATCTCGAGAGAGATGAGAGAGAGAGAGAGATCAACAGATGAGCGCGCGCGAGAGAGAGAG 840
QY 841 AACACCGCGCGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AACACCGCGCGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900

QY 901 TTCCGCGAGTGAACCAAGCGACCCAGGACTTCTTGGAGGTGACAGCTGAGGATCCCCAC 960
DB 901 TTCCGCGAGTGAACCAAGCGACCCAGGACTTCTTGGAGGTGACAGCTGAGGATCCCCAC 960
QY 961 CCCGCGGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
DB 961 CCCGCGGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 AGCGTGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AGCGTGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AACGAGACCCCGGAGATCGGCTTACAGATGACAGAGTGTGCGCGCGAGGCGTGAAGAGGAGC 1140
DB 1081 AACGAGACCCCGGAGATCGGCTTACAGATGACAGAGTGTGCGCGCGAGGCGTGAAGAGGAGC 1140
QY 1141 CCAGCATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 CCAGCATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGATGATCTACAGAGCGCGCGCGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 GAGATGATGATCTACAGAGCGCGCGCGTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GCGCGCAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 GCGCGCAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 TGAAGCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 TGAAGCGTGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 AAGCTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 CTGTGCAAGCTGTGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 CTGTGCAAGCTGTGCGGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCGGAGCTGAGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCGGAGCTGAGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 TACGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 TACGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TACGAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 TACGAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 AGCATGATGATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 AGCATGATGATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 GAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1861 GAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 ACCCGCGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 ACCCGCGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980

QY 1981 ACCCTCTAGTGGAGCGCGCCGCAACCGCGAGACCAAGATCGGGAAGCGCGCTACGG 2040
DB 1981 ACCCTCTAGTGGAGCGCGCCGCAACCGCGAGACCAAGATCGGGAAGCGCGCTACGG 2040
QY 2041 ACCGACCGGGGGCGCGCAAGATCGTGAACCTGACCCGAGACCAACCAAGAAAGCCGAG 2100
DB 2041 ACCGACCGGGGGCGCGCAAGATCGTGAACCTGACCCGAGACCAACCAAGAAAGCCGAG 2100
QY 2101 CTGCAAGCGCATCGAGCTGGCCCTGCAAGACAGCGGAGCGAGGTGAACATCGTGAACGAC 2160
DB 2101 CTGCAAGCGCATCGAGCTGGCCCTGCAAGACAGCGGAGCGAGGTGAACATCGTGAACGAC 2160
QY 2161 ACCGAGTACGCGCTGGGCGATCATCCAGGCGCCGAGCCGCAAGAGCGAGGCGAGCTGGTG 2220
DB 2161 ACCGAGTACGCGCTGGGCGATCATCCAGGCGCCGAGCCGCAAGAGCGAGGCGAGCTGGTG 2220
QY 2221 AACCGATCATCGAGCGCTGATCAAGAGAGAAAGGTGATCCTGAGCTGGGTGCCGCGC 2280
DB 2221 AACCGATCATCGAGCGCTGATCAAGAGAGAAAGGTGATCCTGAGCTGGGTGCCGCGC 2280
QY 2281 CACAAAGGCGATCGCGCGCAACGACGATCGAACAGCTGTGAGCAAGGGCATTCGCGAAG 2340
DB 2281 CACAAAGGCGATCGCGCGCAACGACGATCGAACAGCTGTGAGCAAGGGCATTCGCGAAG 2340
QY 2341 GTGCTGTCTTGAAGCGGATCGATGCGGCGATCGTGTATCAACGATCAATGACACCTG 2400
DB 2341 GTGCTGTCTTGAAGCGGATCGATGCGGCGATCGTGTATCAACGATCAATGACACCTG 2400
QY 2401 TACGTGGGAGCGCGCGCGCTAGGATGATTAAGCTTCCGCGGGCTAGCACCGGTGA 2460
DB 2401 TACGTGGGAGCGCGCGCGCTAGGATGATTAAGCTTCCGCGGGCTAGCACCGGTGA 2460
QY 2461 TTC 2463
DB 2461 TTC 2463

RESULT 2

ADM73765
ID ADM73765 standard; DNA; 2463 BP.

AC ADM73765;

DT 03-JUN-2004 (first entry)

DE HIV-1 polynucleotide #8.

KM HIV-1; gene; de; HIV pol; immune response; DNA immunisation;

KW HIV type C protein; immunostimulant.

OS Human immunodeficiency virus 1.

PN US2003223961-A1.

PD 04-DEC-2003;

PF 05-JUL-2001; 2001US-00899575.

PR 05-JUL-2000; 2000US-00610313.

PA (MEGE/) MEGEDE J Z.

PA (BARN/) BARNETT S W.

PA (ENGE/) ENGELBRECHT S.

PA (RENS/) RENSBURG E J V.

PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;

DR WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV type
PT C proteins.

XX
PS Claim 1; SEQ ID NO 31; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 2463; DB 12; Length 2463;

Best Local Similarity 100.0%; Pred. No. 4,46-297;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGCGGACCATGCGCGAGGCGCATGAGCCGACAGGCGCAATCTGATGCGAG 60
DB 1 GTGAGCGGACCATGCGCGAGGCGCATGAGCCGACAGGCGCAATCTGATGCGAG 60
QY 61 CGCAGCAATCTGAGAGGCGCGCGAGCATCATAGTCTTCACTGCGCGAAGAGGCG 120
DB 61 CGCAGCAATCTGAGAGGCGCGCGAGCATCATAGTCTTCACTGCGCGAAGAGGCG 120
QY 121 CATATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 121 CATATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 GCGCAGCAGATGAAGAGCTGCAACGAGCGCGCGCAACTCTTCCGCGAGGACCTGGCG 240
DB 181 GCGCAGCAGATGAAGAGCTGCAACGAGCGCGCGCAACTCTTCCGCGAGGACCTGGCG 240
QY 241 TTCCCGCGAGGCGAGGCGCGCGAGTTCGCCAGCGAGCAGAACCGCGCGCGCGCGCG 300
DB 241 TTCCCGCGAGGCGAGGCGCGCGAGTTCGCCAGCGAGCAGAACCGCGCGCGCGCGCG 300
QY 301 AGCGCGAGCTGCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 301 AGCGCGAGCTGCAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 GCGCAGCTGCAATCTTCCCGCGAGTCAACCTGTGTGCGAGCGCGCGCGCGCGCGCG 420
DB 361 GCGCAGCTGCAATCTTCCCGCGAGTCAACCTGTGTGCGAGCGCGCGCGCGCGCGCG 420
QY 421 GCGCGCAGATGAAGAGGCGCGCTGTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 421 GCGCGCAGATGAAGAGGCGCGCTGTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 ATGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
DB 481 ATGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 GTGCGCAGTACGACCAAGTCTGTATCGAGTCTGTGCGCGCGCGCGCGCGCGCGCG 600
DB 541 GTGCGCAGTACGACCAAGTCTGTATCGAGTCTGTGCGCGCGCGCGCGCGCGCGCG 600
QY 601 CTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGAATCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 ACCCTGAATCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780

XX (CHTR) CHIRON CORP.
PI Zur Megede J, Barnett SW, Lian Y;
XX WPI: 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
XX The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;
Query Match 99.7%; Score 2455.4; DB 8; Length 2457;
Best Local Similarity 100.0%; Pred. No. 3,9e-296;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTGACGCGCAGCATGCGCGAGGCGCATGAGCCAGGCCAGCGCCAACTCTGATGCG 60
DB 1 GTGACGCGCAGCATGCGCGAGGCGCATGAGCCAGGCCAGCGCCAACTCTGATGCG 60
QY 61 CGCAGCACTTCAAGGGGCCCCCAAGCGCATCATCAAGTCTTCACTGCGCGAAGAGGGC 120
DB 61 CGCAGCACTTCAAGGGGCCCCCAAGCGCATCATCAAGTCTTCACTGCGCGAAGAGGGC 120
QY 121 CACATGCGCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 121 CACATGCGCGCGCATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 GGCACACAGATGAGAGCTGCAACGAGCGCGAGGCCAATCTTCCGCGAGGACTGCGCC 240
DB 181 GGCACACAGATGAGAGCTGCAACGAGCGCGAGGCCAATCTTCCGCGAGGACTGCGCC 240
QY 241 TTCCCGCAGGGGCGAGGCGCGCGAGTTTCCCAAGCGAGAGAGAGCCGCCCAACGCCCCACC 300
DB 241 TTCCCGCAGGGGCGAGGCGCGCGAGTTTCCCAAGCGAGAGAGAGCCGCCCAACGCCCCACC 300
QY 301 AGCGCGAGGTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 301 AGCGCGAGGTGCGAGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 GGCACCTGAACCTTCCCGCGAGTCAACCTGTGCGAGGCGCGCGCGCGCGCGCGCGCG 420
DB 361 GGCACCTGAACCTTCCCGCGAGTCAACCTGTGCGAGGCGCGCGCGCGCGCGCGCGCG 420
QY 421 GCGCGCGAGATCAAGAGGCGCGCGTGTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 421 GCGCGCGAGATCAAGAGGCGCGCGTGTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 ATGAGCTTCCCGCGCGAGTGAAGCCCAAGATGATGCGCGCGCATTCGCGCGCTTCACTAG 540
DB 481 ATGAGCTTCCCGCGCGAGTGAAGCCCAAGATGATGCGCGCGCATTCGCGCGCTTCACTAG 540
QY 541 GTGCGCGAGTCAAGAGCATCTGATGAGATCTGCGCGAGAGAGGCGCATCGCGCGCGT 600
DB 541 GTGCGCGAGTCAAGAGCATCTGATGAGATCTGCGCGAGAGAGGCGCATCGCGCGCGT 600

QY 601 CTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGAACCTTCCCGCGAGTGAAGCCCGCATGAGCCGTCGCGCGCGCGCGCGCGCG 720
DB 661 ACCCTGAACCTTCCCGCGAGTGAAGCCCGCATGAGCCGTCGCGCGCGCGCGCGCGCG 720
QY 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 781 ATTCGCGAGAGATGAGAGAGGCGCGAGATCAACAGATGAGCGCGCGCGCGCGCGCG 840
DB 781 ATTCGCGAGAGATGAGAGAGGCGCGAGATCAACAGATGAGCGCGCGCGCGCGCGCG 840
QY 841 AACACCGCGCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AACACCGCGCGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCCCGCGAGTGAACAGCGCGACCCAGAGCTTCTGCGAGGTGAGCTGCGCGCATTCGCC 960
DB 901 TTCCCGCGAGTGAACAGCGCGACCCAGAGCTTCTGCGAGGTGAGCTGCGCGCATTCGCC 960
QY 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
QY 1021 AGCGTGCCTGCGAGAGAGCTTCCGCGAGTCAACCGCGCTTCAACATCCCGCGAGTCAAC 1080
DB 1021 AGCGTGCCTGCGAGAGAGCTTCCGCGAGTCAACCGCGCTTCAACATCCCGCGAGTCAAC 1080
QY 1081 AACGAGACCGCGCGCATCCGCTACAGATCAAGTGTGCGCGCGCGCGCGCGCGCGCG 1140
DB 1081 AACGAGACCGCGCGCATCCGCTACAGATCAAGTGTGCGCGCGCGCGCGCGCGCGCG 1140
QY 1141 CCAGCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 CCAGCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGATCTACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
DB 1201 GAGATGATCTACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1261 CGCGCGAGATGAGAGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 CGCGCGAGATGAGAGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGACAGAGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
DB 1321 AAGAGACAGAGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 TGAACCGTGAAGCGCATGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 TGAACCGTGAAGCGCATGAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AAGCTGTGTGCGAGAGTGAAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 AAGCTGTGTGCGAGAGTGAAGTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 CTGTGCGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
DB 1501 CTGTGCGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
QY 1561 GCGGAGCTGAGAGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCGGAGCTGAGAGTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 TACGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 TACGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TACGAGATCTACAGAGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

Db 1681 TACGAGATCTACCAAGAGCCCTTCAAGAACTGAAAGACCGGCAAGTACGCCAAGATGCGC 1740
QY 1741 ACCGCCACACCAAGAGAGCTGAAAGACGCTGACCGAGCGCTGCAAGATGCGC 1800
Db 1741 ACCGCCACACCAAGAGAGCTGAAAGACGCTGACCGAGCGCTGCAAGATGCGC 1800
QY 1801 AGCATGTAATCTGAGGAGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGACTGAG 1860
Db 1801 AGCATGTAATCTGAGGAGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGACTGAG 1860
QY 1861 GAGACCTGTGAGACGACTACTGAGACGCGACCTGAGATCCCGAGTGGAGTTCTGTAAC 1920
Db 1861 GAGACCTGTGAGACGACTACTGAGACGCGACCTGAGATCCCGAGTGGAGTTCTGTAAC 1920
QY 1921 ACCCCCCCTGTAAGTGTGTAACGAGCTGAGAGAGAGACCCATCTCCGCGCCGAG 1980
Db 1921 ACCCCCCCTGTAAGTGTGTAACGAGCTGAGAGAGAGACCCATCTCCGCGCCGAG 1980
QY 1981 ACCTTCTAGTGAAGCGCGCCGCAACCGAGAGACCAAGATGAGGACCGGCTAAGTG 2040
Db 1981 ACCTTCTAGTGAAGCGCGCCGCAACCGAGAGACCAAGATGAGGACCGGCTAAGTG 2040
QY 2041 ACCGACCGGAGCGCGAGAAAGTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGAGCGCGAGAAAGTGTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
QY 2101 CTGAGAGGCAATCAAGTGTGAGCTGAGAGACGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 CTGAGAGGCAATCAAGTGTGAGCTGAGAGACGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 AGCCAGTACGCGCTGAGAGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 AGCCAGTACGCGCTGAGAGCAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 AACGAGTATGAGAGAGTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db 2221 AACGAGTATGAGAGAGTGTGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 CACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2341 GTGCTGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db 2341 GTGCTGTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
QY 2401 TACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2457
Db 2401 TACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2457

RESULT 4
ADCL3265
ID ADCL3265 standard; DNA; 2457 BP.

AC ADCL3265;
DT 18-DEC-2003 (first entry)
DE DNA of HIV construct p2Pol-opt-ym_C SEQ ID NO 44.
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
OS Human immunodeficiency virus.
PN WO2003004620-A2.
XX 16-JAN-2003.
PD 05-JUL-2002; 2002WO-US021420.
XX

PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
PI Zur Megele J, Barnette SW, Llan Y, Engelbrecht S, Van Rensburg RJ,
XX MPI; 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX PT Prot, or Rev polypeptide, useful for immunisation, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 41; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 99.7%; Score 2455.4; DB 10; Length 2457;
Blast Local Similarity 100.0%; Pred. No. 3.9e-296;

Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTGAGAGCAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
Db 1 GTGAGAGCAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 61 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 61 CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 CAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db 181 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 TTCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Db 241 TTCCCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 301 AGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db 301 AGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 361 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 421 GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 ATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Db 481 ATGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 GTGAGAGCAGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

Db 541 GTGCGCAGTACGACCAAGATCTTGATCGAGATCTGGGCGAAGAGGCCATCGGACCGTG 600
 Qy 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCGACATGCTGACCCGACGTGGCTGC 660
 Db 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGCGACATGCTGACCCGACGTGGCTGC 660
 Qy 661 ACCCTGAATTTCCCATCGAGCCCATCGAGACCTGTGACCCGTGAAAGCTGAAAGCCCGGCA 720
 Db 661 ACCCTGAATTTCCCATCGAGCCCATCGAGACCTGTGACCCGTGAAAGCTGAAAGCCCGGCA 720
 Qy 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCCGAGAGAAATCAAGGCCCTGAACCGGC 780
 Db 721 GACGCGCCCAAGGTGAAGAGTGGCCCTGACCCGAGAGAAATCAAGGCCCTGAACCGGC 780
 Qy 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTTAC 840
 Db 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCCGAGAACCCCTTAC 840
 Qy 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGTGAC 900
 Db 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGTGAC 900
 Qy 901 TTCCGAGACTGAACAAGCGCACCCGAGACTTCTGGAGGTGACGCTGGGCAATCCGCCAC 960
 Db 901 TTCCGAGACTGAACAAGCGCACCCGAGACTTCTGGAGGTGACGCTGGGCAATCCGCCAC 960
 Qy 961 CCGCGCGGCTGAAAGAGAGAGACGCTGACCGTCTGACCGTGGCGACCGCTTACTTC 1020
 Db 961 CCGCGCGGCTGAAAGAGAGAGACGCTGACCGTCTGACCGTGGCGACCGCTTACTTC 1020
 Qy 1021 AGCGTGCCTCTGAGCAAGAGCTTCCGCAAGTACACGCTTCAACATCCCAAGATCAAC 1080
 Db 1021 AGCGTGCCTCTGAGCAAGAGCTTCCGCAAGTACACGCTTCAACATCCCAAGATCAAC 1080
 Qy 1081 AACGAGACCCCGGCGCATCCGCTACCAAGTACCAAGTCTGCGCCAGGCTGAAAGGCGAC 1140
 Db 1081 AACGAGACCCCGGCGCATCCGCTACCAAGTACCAAGTCTGCGCCAGGCTGAAAGGCGAC 1140
 Qy 1141 CCGAGCATCTTCCAGAGACGATAGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCC 1200
 Db 1141 CCGAGCATCTTCCAGAGACGATAGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCC 1200
 Qy 1201 GAGATGTGATCTACCAAGGCGCCCTGTACGTGGGACGACCTGAGATCGGACGACAC 1260
 Db 1201 GAGATGTGATCTACCAAGGCGCCCTGTACGTGGGACGACCTGAGATCGGACGACAC 1260
 Qy 1261 CCGGCGCAAGTCAAGAGAGCTGCGACACCTGCTGGCTGGGGCTTCAACACCCCGAC 1320
 Db 1261 CCGGCGCAAGTCAAGAGAGCTGCGACACCTGCTGGCTGGGGCTTCAACACCCCGAC 1320
 Qy 1321 AAGAGACACCAAGAGAGGCCCCCTTCTGTGATGGGCTACAGAGCTGACCCCGACAG 1380
 Db 1321 AAGAGACACCAAGAGAGGCCCCCTTCTGTGATGGGCTACAGAGCTGACCCCGACAG 1380
 Qy 1381 TGGACCGTGAAGCCCATCGAGCTGCGCGAGAGAGAGCTGACCGTGAACGACATTCAG 1440
 Db 1381 TGGACCGTGAAGCCCATCGAGCTGCGCGAGAGAGAGCTGACCGTGAACGACATTCAG 1440
 Qy 1441 AAGCTGTGGGCAAGCTGAACTGGGCGACGACATCTAACCCCGGCAATCAAGGTGGCGAG 1500
 Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGACGACATCTAACCCCGGCAATCAAGGTGGCGAG 1500
 Qy 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCTGACCGAGAG 1560
 Db 1501 CTGTGCAAGCTGTGCGCGCGCGCGCAAGGCGCTGACCGACATCTGTGCCCTGACCGAGAG 1560
 Qy 1561 GCGGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGCGAGCCCGTGCACCGCGGTGAC 1620
 Db 1561 GCGGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGCGAGCCCGTGCACCGCGGTGAC 1620
 Qy 1621 TACGACCCCGAGAGAGCTGTGTGGCCGAGATCCAGAAAGAGGGCCACGACATGTGACC 1680
 Db 1621 TACGACCCCGAGAGAGCTGTGTGGCCGAGATCCAGAAAGAGGGCCACGACATGTGACC 1680

Db 1621 TACGACCCCGAGAGAGCTGTGTGGCCGAGATCCAGAAAGAGGGCCACGACATGTGACC 1680
 Qy 1681 TACGAGATCTACAGAGAGGCTTCAAGAACTTGAAAGCCGGAAGTAACCCCAAGATGCC 1740
 Db 1681 TACGAGATCTACAGAGAGGCTTCAAGAACTTGAAAGCCGGAAGTAACCCCAAGATGCC 1740
 Qy 1741 ACCGCGCACCAAGAGAGAGCTGAAAGAGCTGACCGAGGCGGTGACAGAAATGCCATGAG 1800
 Db 1741 ACCGCGCACCAAGAGAGAGCTGAAAGAGCTGACCGAGGCGGTGACAGAAATGCCATGAG 1800
 Qy 1801 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTGCGCATCCAGAGAGAGACTGG 1860
 Db 1801 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCCGCTGCGCATCCAGAGAGAGACTGG 1860
 Qy 1861 GAGACCTGTGAGCCGACTACTGAGCGGCGACCTGATCCCGAGTGGGAGTTCTGTAAC 1920
 Db 1861 GAGACCTGTGAGCCGACTACTGAGCGGCGACCTGATCCCGAGTGGGAGTTCTGTAAC 1920
 Qy 1921 ACCCGCGCGGTGAGAGCTGTGTGTAACGAGCTGAGAGAGAGAGGCCATCATCGGCGCGAG 1980
 Db 1921 ACCCGCGCGGTGAGAGCTGTGTGTAACGAGCTGAGAGAGAGAGGCCATCATCGGCGCGAG 1980
 Qy 1981 ACCCTTACGTGAGACGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTG 2040
 Db 1981 ACCCTTACGTGAGACGCGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACGTG 2040
 Qy 2041 ACCGACCGGCGCGCGAGAGATCTGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100
 Db 2041 ACCGACCGGCGCGCGAGAGATCTGTAGCTGACCGAGACCAACCAAGAGACCGAG 2100
 Qy 2101 CTGCGAGCCCATCAAGTGGCCCTGCGAGGACGCGGCGAGAGGTTGAATCTGTGACCGAC 2160
 Db 2101 CTGCGAGCCCATCAAGTGGCCCTGCGAGGACGCGGCGAGAGGTTGAATCTGTGACCGAC 2160
 Qy 2161 AGCAGTACGCGCTGTGGCATCATCCAGGCCACGCCGCAAGAGCGAGCGACTGTG 2220
 Db 2161 AGCAGTACGCGCTGTGGCATCATCCAGGCCACGCCGCAAGAGCGAGCGACTGTG 2220
 Qy 2221 AACGAGTATGAGAGAGCTGATCAAGAGAGAGAGGTGTACTTGAGCTGTGGTCCGCC 2280
 Db 2221 AACGAGTATGAGAGAGCTGATCAAGAGAGAGAGGTGTACTTGAGCTGTGGTCCGCC 2280
 Qy 2281 CACAGAGGATCGCGCGGCAACGAGAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAG 2340
 Db 2281 CACAGAGGATCGCGCGGCAACGAGAGATCGACAAGCTGTGAGCAAGGGCATCCGCAAG 2340
 Qy 2341 GTGCTGTTCTGTGAGCGGATCGATGCGGCGCATGTGATCTACCAAGTACATGAGACGACTG 2400
 Db 2341 GTGCTGTTCTGTGAGCGGATCGATGCGGCGCATGTGATCTACCAAGTACATGAGACGACTG 2400
 Qy 2401 TACGTGGGCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCTAGACCGGT 2457
 Db 2401 TACGTGGGCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCTAGACCGGT 2457
 RESULT 5
 ABLJ39959
 ID ABLJ39959 standard; DNA; 2469 BP.
 XX
 AC ABLJ39959;
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Synthetic construct PR975(+) SEQ ID NO:30.
 XX
 KW Human immunodeficiency virus type C; antigenic HIV type C protein;
 KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;
 KW immunostimulant; gene therapy; gene; de.
 XX
 OS Human immunodeficiency virus; type C.
 OS Synthetic.
 XX
 PN W0200204493-A2.

Db 361 GGACACCTTGAACTTCCCCCAGATCACTCTGTGGACAGCCCTCTGTGATGATCAAGTGG 420
Qy 421 GGCGGCGAGATCAAGAGAGGCGCTGTGACACCGAGGCGAGACACCGTGTGAGAG 480
Db 421 GGCGGCGAGATCAAGAGAGGCGCTGTGACACCGAGGCGAGACACCGTGTGAGAG 480
Qy 481 ATGAGCCTTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGAGCATCGGCGCTTCAACAG 540
Db 481 ATGAGCCTTGCCCGGCAAGTGAAGCCCAAGATGATCGGCGAGCATCGGCGCTTCAACAG 540
Qy 541 GTGGCCCGATGATCAACATCTCTGATCGAGATCTGCGGAGAGAGAGCCATCGGCACTG 600
Db 541 GTGGCCCGATGATCAACATCTCTGATCGAGATCTGCGGAGAGAGAGCCATCGGCACTG 600
Qy 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGGCAACATCTGATCCCACTGGAGCTGC 660
Db 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGGCAACATCTGATCCCACTGGAGCTGC 660
Qy 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGCAATG 720
Db 661 ACCCTGAATCTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAAGCCCGCAATG 720
Qy 721 GACGCGCCCCAAGGTGAAGAGTGGGCGCTGACCGAGAGAGAGATCAAGGCGCTGACCGCC 780
Db 721 GACGCGCCCCAAGGTGAAGAGTGGGCGCTGACCGAGAGAGAGATCAAGGCGCTGACCGCC 780
Qy 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGGAGAACCCCTTAC 840
Db 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGGAGAACCCCTTAC 840
Qy 841 AACACCCCGGTGTGGCATCAAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGTGAC 900
Db 841 AACACCCCGGTGTGGCATCAAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGTGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGGAGGTGCAAGTGGGATCTCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGGAGGTGCAAGTGGGATCTCCCCAC 960
Qy 961 CCCGCGCGCTGAAGAGAGAGAGCGGTGACCGTGTGAGAGTGGGCGGACGCTTACTTC 1020
Db 961 CCCGCGCGCTGAAGAGAGAGAGCGGTGACCGTGTGAGAGTGGGCGGACGCTTACTTC 1020
Qy 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCGACATCAAC 1080
Db 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCGACATCAAC 1080
Qy 1081 AACGAGACCCCGGCAATCGGCTTACCAAGTACAGTGTGCTGCCAGGGCTGAGAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCAATCGGCTTACCAAGTACAGTGTGCTGCCAGGGCTGAGAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGAGAGCCCTTCCGCGCCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGAGAGCCCTTCCGCGCCCGCAACCCC 1200
Qy 1201 GAGATCGTATTTACCA-----GGCCCCCTGTACGTGGGAGCGACCTTGAAGATCGAC 1254
Db 1201 GAGATCGTATTTACCA-----GGCCCCCTGTACGTGGGAGCGACCTTGAAGATCGAC 1254
Qy 1255 CAGGACCGGCGCAAGTGAAGAGTGTGCGCAAGACCTGTGCGCTGGGGCTTCAACACCC 1314
Db 1255 CAGGACCGGCGCAAGTGAAGAGTGTGCGCAAGACCTGTGCGCTGGGGCTTCAACACCC 1314
Qy 1261 CAGGACCGGCGCAAGTGAAGAGTGTGCGCAAGACCTGTGCGCTGGGGCTTCAACACCC 1320
Db 1261 CAGGACCGGCGCAAGTGAAGAGTGTGCGCAAGACCTGTGCGCTGGGGCTTCAACACCC 1320
Qy 1315 CCCGACAAAGAGACCAAGAGAGGCCCCCTTCTGTGATGAGGCTTCAAGGCTGACCCC 1374
Db 1315 CCCGACAAAGAGACCAAGAGAGGCCCCCTTCTGTGATGAGGCTTCAAGGCTGACCCC 1374
Qy 1375 GACAAAGTGAACCGTGAAGCCCATTCGAGCTGCCGAGAGAGAGAGTGAACCGTGAACGAC 1434
Db 1375 GACAAAGTGAACCGTGAAGCCCATTCGAGCTGCCGAGAGAGAGAGTGAACCGTGAACGAC 1434
Qy 1435 ATCCGAAGAGCTGTGGGCAAGCTGAATCTGGGCCAAGCCAGATCTTACCCCGGCAATCAAGTGG 1494
Db 1435 ATCCGAAGAGCTGTGGGCAAGCTGAATCTGGGCCAAGCCAGATCTTACCCCGGCAATCAAGTGG 1494
Qy 1441 ATCCGAAGAGCTGTGGGCAAGCTGAATCTGGGCCAAGCCAGATCTTACCCCGGCAATCAAGTGG 1500
Db 1441 ATCCGAAGAGCTGTGGGCAAGCTGAATCTGGGCCAAGCCAGATCTTACCCCGGCAATCAAGTGG 1500

Qy 1495 CGCCAGCTGTGAAGCTGTGGCGGCGCCAGAGGCGCTGACCGACATGTGCGCCCTGACCC 1554
Db 1501 CGCCAGCTGTGAAGCTGTGGCGGCGCCAGAGGCGCTGACCGACATGTGCGCCCTGACCC 1560
Qy 1555 GAGAGGCGAGAGCTGTGAAGCTGTGGCGGAGAACCGGAGATCTTGGCGGAGCCCGTGAACG 1614
Db 1561 GAGAGGCGAGAGCTGTGAAGCTGTGGCGGAGAACCGGAGATCTTGGCGGAGCCCGTGAACG 1620
Qy 1615 GTGTACTAGACCCCGAGAGAGAGCTGTGGCGGAGATCTTGAAGAGAGAGAGAGAGAGAG 1674
Db 1621 GTGTACTAGACCCCGAGAGAGAGCTGTGGCGGAGATCTTGAAGAGAGAGAGAGAGAGAG 1680
Qy 1675 TGAACCTACAGATCTTACAGAGAGGCGCTTCAAGAACTGAAAGACCGGAGAGAGAGAGAG 1734
Db 1681 TGAACCTACAGATCTTACAGAGAGGCGCTTCAAGAACTGAAAGAGAGAGAGAGAGAGAG 1740
Qy 1735 ATGCGACCGCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Db 1741 ATGCGACCGCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1795 ATGAGAGAGATCTGTATCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Db 1801 ATGAGAGAGATCTGTATCTGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1855 ACTGAGAGAGAGCTGTGTGAACCGACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Db 1861 ACTGAGAGAGAGCTGTGTGAACCGACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy 1915 GTGAACACCCCGCTGTGTGAAGCTGTGTGAACAGCTGAGAGAGAGAGAGAGAGAGAG 1974
Db 1921 GTGAACACCCCGCTGTGTGAAGCTGTGTGAACAGCTGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1975 GCGGAGAGAGCTTCTAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
Db 1981 GCGGAGAGAGCTTCTAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy 2035 TACGTGAACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
Db 2041 TACGTGAACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2095 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
Db 2101 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy 2155 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
Db 2161 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2215 CTGTGTGAACAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
Db 2221 CTGTGTGAACAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy 2275 CCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
Db 2281 CCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2335 CGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
Db 2341 CGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy 2395 GACCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
Db 2401 GACCTGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469

RESULT 7
ABI3961
ID ABI3961 standard; DNA; 2457 BP.

XX ABL39961;
 AC
 XX
 DT 15-MAY-2002 (first entry)
 XX
 DE Synthetic construct PR975YMMN SEQ ID NO:32.
 XX
 KW Human immunodeficiency virus type C; antigenic HIV type C protein;
 KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
 KW immunostimulant; gene therapy; gene; da.
 XX
 OS Human immunodeficiency virus; type C.
 OS Synthetic.
 PN WO200204493-A2.
 XX
 PD 17-JAN-2002.
 XX
 PF 05-JUL-2001; 2001WO-US021241.
 XX
 PR 05-JUL-2000; 2000US-00610313.
 XX
 PA (CHIR) CHIRON CORP.
 PA (UYST-) UNIV STELLENBOSCH.
 XX
 PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
 DR WPI; 2002-154920/20.
 XX
 PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
 PT applications including DNA immunization or generation of packaging cell
 PT lines, particularly in gene therapy.
 XX
 PS Claim 1; Fig 10; 233pp; English.
 XX
 CC The present invention describes expression cassettes comprising a
 CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
 CC type C polypeptides. The expression cassettes comprise any of the HIV
 CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
 CC (1). (i) have immunostimulant activity and can be used in gene therapy.
 CC The HIV type C polynucleotides are useful in applications including DNA
 CC immunisation, generation of packaging cell lines, and production of HIV
 CC type C proteins. The polynucleotides are particularly useful in gene
 CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
 CC ABB06204 to ABB06215 represent sequences used in the exemplification of
 CC the present invention
 CC
 SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other:
 Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
 Best Local Similarity 99.6%; Pred. No. 9.3e-294;
 Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 301 AGCCGAGCTGCAGGTGCGCGGCGAACACCCCGCAGCGAGCCGCGCCGCGCAG 360
 DB |||||
 QY 301 AGCCGAGCTGCAGGTGCGCGGCGAACACCCCGCAGCGAGCCGCGCCGCGCAG 360
 DB |||||
 QY 361 GGCACCTTGAATTTCCCGCAATCACTGTGAGCGCGCCCTGTGTAGCATCAAGTGTG 420
 DB |||||
 QY 361 GGCACCTTGAATTTCCCGCAATCACTGTGAGCGCGCCCTGTGTAGCATCAAGTGTG 420
 DB |||||
 QY 421 GCGCGCAGATCAAGAGAGCCCTGTGTAGCACTCGCGCGCGCAACACCTGTGTAGAGAG 480
 DB |||||
 QY 421 GCGCGCAGATCAAGAGAGCCCTGTGTAGCACTCGCGCGCGCAACACCTGTGTAGAGAG 480
 DB |||||
 QY 481 ATGAGCTGCGCGGCAATGGAAGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
 DB |||||
 QY 481 ATGAGCTGCGCGGCAATGGAAGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
 DB |||||
 QY 541 GTGCGCAGTACGACAGATCTGTATCGATCTGCGCGCAAGGCCATCGGCACTGTG 600
 DB |||||
 QY 541 GTGCGCAGTACGACAGATCTGTATCGATCTGCGCGCAAGGCCATCGGCACTGTG 600
 DB |||||
 QY 601 CTGATCGGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATCTGACCTGAGCTTGC 660
 DB |||||
 QY 601 CTGATCGGCGCCCAACCCCGGTGAACATCATCGCGCGCAACATCTGACCTGAGCTTGC 660
 DB |||||
 QY 661 ACCCTGAATCTCCCATACAGCCCATCGAGACCGGTGCGGTGAGCTGAAAGCCCGCATG 720
 DB |||||
 QY 661 ACCCTGAATCTCCCATACAGCCCATCGAGACCGGTGCGGTGAGCTGAAAGCCCGCATG 720
 DB |||||
 QY 721 GACGCGCCCAAGGTGAAGAGTGTGCGCCCTGACCGGAGAGAAATCAAGCCCTGACCGCC 780
 DB |||||
 QY 721 GACGCGCCCAAGGTGAAGAGTGTGCGCCCTGACCGGAGAGAAATCAAGCCCTGACCGCC 780
 DB |||||
 QY 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACAGATGTGCGCCGAGAACCTCTAC 840
 DB |||||
 QY 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACAGATGTGCGCCGAGAACCTCTAC 840
 DB |||||
 QY 841 AACACCCCGTGTGCGCATCAAGAGAGAGCAAGCAAGTGTGCGGAGCGTGTGAGAC 900
 DB |||||
 QY 841 AACACCCCGTGTGCGCATCAAGAGAGAGCAAGCAAGTGTGCGGAGCGTGTGAGAC 900
 DB |||||
 QY 901 TTCGCGAGCTGAACAAGGCGCACCCGAGACTTCTGTGAGGTGAGCTGTGAGATCCCGCAC 960
 DB |||||
 QY 901 TTCGCGAGCTGAACAAGGCGCACCCGAGACTTCTGTGAGGTGAGCTGTGAGATCCCGCAC 960
 DB |||||
 QY 961 CCCGCGGCTTGAAGAGAGAGAGAGCGTGTGACCGTGTGAGCGTGTGAGCGCATCTTC 1020
 DB |||||
 QY 961 CCCGCGGCTTGAAGAGAGAGAGAGCGTGTGACCGTGTGAGCGTGTGAGCGCATCTTC 1020
 DB |||||
 QY 1021 AGCGTGTGCTGTGAGAGAGATCTTCCGCAAGTACACCGCTTCAATCCCGCAAGATCAAC 1080
 DB |||||
 QY 1021 AGCGTGTGCTGTGAGAGAGATCTTCCGCAAGTACACCGCTTCAATCCCGCAAGATCAAC 1080
 DB |||||
 QY 1081 AACGAGAGCCCGGCGCATCGCTACAGTACCAAGTGTGAGCGTGTGAGAGGCGCAGC 1140
 DB |||||
 QY 1081 AACGAGAGCCCGGCGCATCGCTACAGTACCAAGTGTGAGCGTGTGAGAGGCGCAGC 1140
 DB |||||
 QY 1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCCGCAACCC 1200
 DB |||||
 QY 1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCCGCAACCC 1200
 DB |||||
 QY 1201 GAGATGTGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 DB |||||
 QY 1201 GAGATGTGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 DB |||||
 QY 1261 CGGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 DB |||||
 QY 1261 CGGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 DB |||||
 QY 1321 AAGAGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 DB |||||
 QY 1321 AAGAGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 DB |||||
 QY 1321 AAGAGACACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
 DB |||||
 QY 1381 TGAACCGTGAAGCCCATCGAGCTGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 DB |||||

```
Db 1375 TGGACCGTGACAGCCATCGAGCTGCCCCGAGAGAGAGCTGACCGTGAACGACATCCAG 1434
1441 AAGCTGTGTGGGAGAGCTGAAGTGGGACACGAGATCTACCCCGGATCGAAGGTGGCCAG 1500
1435 AAGCTGTGTGGGAGAGCTGAAGTGGGACACGAGATCTACCCCGGATCGAAGGTGGCCAG 1494
1501 CTGTGCAAGCTGTGTGGGACCGGACCAAGGCTTGAACGACATGTGTGCTTGAACCGAGAG 1560
1495 CTGTGCAAGCTGTGTGGGACCGGACCAAGGCTTGAACGACATGTGTGCTTGAACCGAGAG 1554
1561 GCCGAGCTGTGTGGGACCGGACCAAGGATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
1555 GCCGAGCTGTGTGGGACCGGACCAAGGATCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1614
1621 TACGACCCGAGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
1615 TACGACCCGAGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1674
1681 TACGAGATCTTACGAGAGAGCTTTCAGAGACTTGAAGACCGGACAGTACCGCAAGATGCGC 1740
1675 TACGAGATCTTACGAGAGAGCTTTCAGAGACTTGAAGACCGGACAGTACCGCAAGATGCGC 1734
1741 ACCGCCCAACCAACGACGTGAAGCAGTGAACCGAGAGCGTGTGCAAGATGCGCAATGAG 1800
1735 ACCGCCCAACCAACGACGTGAAGCAGTGAACCGAGAGCGTGTGCAAGATGCGCAATGAG 1794
1801 AACATCGTGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
1795 AACATCGTGTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1854
1861 GAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
1855 GAGACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1914
1921 ACCCCCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
1915 ACCCCCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1974
1981 ACCCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
1975 ACCCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2034
2041 ACCGACCGGAGGCGGAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGCCAG 2100
2035 ACCGACCGGAGGCGGAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGCCAG 2094
2101 CTGCAAGGCGCATCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
2095 CTGCAAGGCGCATCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2154
2161 ACCGAGTACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
2155 ACCGAGTACGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2214
2221 AACGAGATCATCGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280
2215 AACGAGATCATCGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2274
2281 CACAAGGAGATCGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
2275 CACAAGGAGATCGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2334
2341 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400
2335 GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2394
2401 TACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2460
2395 TACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2454
2461 TTC 2463
|||
```

```
Db 2455 TTC 2457
RESULT 8
ADM73766
ID ADM73766 standard; DNA; 2457 BP.
XX
AC ADM73766;
XX
DT 03-JUN-2004 (first entry)
XX
DE HIV-1 polynucleotide #9.
XX
KW HIV-1; gene; de; HIV pol; immune response; DNA immunisation;
KW HIV type C protein; immunostimulant.
XX
OS Human immunodeficiency virus 1.
XX
PN US2003233961-A1.
XX
PD 04-DEC-2003.
XX
PF 05-JUL-2001; 2001US-00899575.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (MEGR/) MEGREDE J Z.
PA (BARN/) BARNETT S W.
PA (ENG/) ENGELBRECHT S.
PA (RENS/) RENSBERG B J V.
XX
PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;
XX
DR WPI; 2004-060515/06.
XX
PT New expression cassette comprising a polynucleotide sequence encoding an
PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT immunisation, generating of packaging cell lines or in producing HIV Type
PT C proteins.
XX
PS Claim 1; SEQ ID NO 32; 160pp; English.
XX
CC The invention relates to an expression cassette comprising a
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC also relates to a recombinant expression system for use in a host cell
CC comprising an expression cassette, where the polynucleotide sequence
CC further comprises control elements capable of driving expression in the
CC selected host cell, a cell comprising an expression cassette where the
CC polynucleotide sequence further comprises control elements compatible
CC with the expression in the cell and a composition for generating an
CC immunological response, comprising an expression cassette. The expression
CC cassette and the methods of the invention are useful in eliciting an
CC immune response, in DNA immunisation, in generation of packaging cell
CC lines and in producing HIV Type C proteins. This sequence represents an
CC HIV-1 polynucleotide of the invention.
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;
Query Match 98.9%; Score 2436.2; DB 12; Length 2457;
Best Local Similarity 99.6%; Pred. No. 9.3e-294;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGACGCGCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 60
DB 1 GTGACGCGCACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 60
QY 61 CGCAGCACTTCAAGGAGCCCAAGGCGCATCATCAAGTGTCTTCACTGTGCGCAAGAGGAG 120
DB 61 CGCAGCACTTCAAGGAGCCCAAGGCGCATCATCAAGTGTCTTCACTGTGCGCAAGAGGAG 120
QY 121 CACATGCGCGCGCACTGTGCGCGCGCGCGCAAGAGAGGCTGTGTGAAGTGTGCGCAAGAG 180
DB 121 CACATGCGCGCGCACTGTGCGCGCGCGCGCAAGAGAGGCTGTGTGAAGTGTGCGCAAGAG 180
```

QY 181 GGCACAGATGAGAGCTGCACCGAGCCCAAGCTTCTTCCGCGAGAGCTGGCC 240
DB 181 GGCACAGATGAGAGCTGCACCGAGCCCAAGCTTCTTCCGCGAGAGCTGGCC 240
QY 241 TTTCCCGAGGGCAAGGCGCGAGTTCCCGAGAGCAGAACCGGCGCAAGCCCCCACC 300
DB 241 TTTCCCGAGGGCAAGGCGCGAGTTCCCGAGAGCAGAACCGGCGCAAGCCCCCACC 300
QY 301 AACCGGAGCTGCAGGTGCGGCGCAAAACCCCGAGGCGCGCGCGAGCGCCAG 360
DB 301 AACCGGAGCTGCAGGTGCGGCGCAAAACCCCGAGGCGCGCGCGAGCGCCAG 360
QY 361 GGCACCTGAATCTTCCCGAGATCACTCTGTGAGCGGCCCTGTGTAGCATCAAGTG 420
DB 361 GGCACCTGAATCTTCCCGAGATCACTCTGTGAGCGGCCCTGTGTAGCATCAAGTG 420
QY 421 GGCAGCAGATCAAGAGGCGCTGTGAGCAACCGGCGCGAGCGACCGTGTGAGGAG 480
DB 421 GGCAGCAGATCAAGAGGCGCTGTGAGCAACCGGCGCGAGCGACCGTGTGAGGAG 480
QY 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCATCGCGCTTCATCAAG 540
DB 481 ATGAGCTGCGCGCAAGTGAAGCCCAAGATGATGCGCGCATCGCGCTTCATCAAG 540
QY 541 GTGCGCAGATCAAGATCTGTATGAGATCTGTGCGAGAGAGCCATCGGACCTTG 600
DB 541 GTGCGCAGATCAAGATCTGTATGAGATCTGTGCGAGAGAGCCATCGGACCTTG 600
QY 601 CTGATGGGCCCCCCCCCTGGAATCATATGCGCGCAATGTGTGAACCGAGCTGGCTG 660
DB 601 CTGATGGGCCCCCCCCCTGGAATCATATGCGCGCAATGTGTGAACCGAGCTGGCTG 660
QY 661 ACCCTGAATCTTCCCGATCAGCCCATCAGACCGGTGCGGTGAAGCTGAAGCCCGAGT 720
DB 661 ACCCTGAATCTTCCCGATCAGCCCATCAGACCGGTGCGGTGAAGCTGAAGCCCGAGT 720
QY 721 GAGCGCCCAAGGTGAAGAGTGTGCGCTGTGACGAGAGAGATCAAGGCTGTGACGCG 780
DB 721 GAGCGCCCAAGGTGAAGAGTGTGCGCTGTGACGAGAGAGATCAAGGCTGTGACGCG 780
QY 781 ATCTGGAAGAGATGAGAGAGAGGCGAGATCAACCAAGATCGGCGCGAGAACCTCTAC 840
DB 781 ATCTGGAAGAGATGAGAGAGAGGCGAGATCAACCAAGATCGGCGCGAGAACCTCTAC 840
QY 841 AACACCCCGTGTTCGCAATCAAGAGAGAGAGCAAGTGTGCGCAAGCTGTGAGAC 900
DB 841 AACACCCCGTGTTCGCAATCAAGAGAGAGAGCAAGTGTGCGCAAGCTGTGAGAC 900
QY 901 TTCGCGAGCTGAACAAGCGCAACCTTCTGTGAGGTGAGCTGTGAGATCCCCAC 960
DB 901 TTCGCGAGCTGAACAAGCGCAACCTTCTGTGAGGTGAGCTGTGAGATCCCCAC 960
QY 961 CCGCGCGGCTGAAGAGAGAGAGGCTGACCTGTGAGCGTGGGCGACGCTTACTTC 1020
DB 961 CCGCGCGGCTGAAGAGAGAGAGGCTGACCTGTGAGCGTGGGCGACGCTTACTTC 1020
QY 1021 AGGTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AGGTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 CCGAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 CCGAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 GAGATGATGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260

QY 1261 GCGCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 GCGCCAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 TGAACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 TGAACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 AAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1495 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
DB 1495 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1554
QY 1561 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 ACCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 ACCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1795 AGCATGTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
DB 1795 AGCATGTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
QY 1854 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1854 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 ACCCGCGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 ACCCGCGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 ACCCTTCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1981 ACCCTTCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 2041 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
DB 2161 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2221 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
DB 2281 CACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
DB 2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400


```

Db 1081 ACCCGGGGATCCGCTACCAATGACAAAGTGTGCCCCGAGGGGTGAGAGGCGACCCGACG 1140
Qy 1147 ATCTTCCAGAGCAGATGACCAAGATCTGTAGAGCCCTTCCGCGCCCGCAACCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTGTAGAGCCCTTCCGCGCCCGCAACCCGAGATC 1200
Qy 1207 GTGAGTACCA-----GGCCCCCTGTATCGTGGGCAAGGACCTGAGATGAGGCGCAGAC 1260
Db 1201 GTGAGTACCAATGACATGACCACTGTATCGTGGGCAAGGACCTGAGATGAGGCGCAGAC 1260
Qy 1261 CGCGCAAGATGAGGAGCTGTGCAAGCACTGTGCGTGGGGCTTCAACACCCCGAC 1320
Db 1261 CGCGCAAGATGAGGAGCTGTGCAAGCACTGTGCGTGGGGCTTCAACACCCCGAC 1320
Qy 1321 AAGAACACACAGAGAGAGCCCTTCTGTGATGAGGCTACGAGCTGACCCGACAG 1380
Db 1321 AAGAACACACAGAGAGAGCCCTTCTGTGATGAGGCTACGAGCTGACCCGACAG 1380
Qy 1381 TGGACCTGTGACACCCATCGAGCTGCGCGGAGAGAGAGCTGAGACCTGACACATCCAG 1440
Db 1381 TGGACCTGTGACACCCATCGAGCTGCGCGGAGAGAGAGCTGAGACCTGACACATCCAG 1440
Qy 1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGCAGATCTACCCGCGCATCAAGGTGGCGCAG 1500
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGCAGATCTACCCGCGCATCAAGGTGGCGCAG 1500
Qy 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCTGTGACGACATGTGCGCTGTGACCGAGAG 1560
Db 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCTGTGACGACATGTGCGCTGTGACCGAGAG 1560
Qy 1561 GCGGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGTGCGGAGCCGTGCACGCGTGTAC 1620
Db 1561 GCGGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGTGCGGAGCCGTGCACGCGTGTAC 1620
Qy 1621 TACGACCCCGAGAGAGAGCTGTGCGCGAGATCTGAGAACGAGGCGCACAGCATGTGAGC 1680
Db 1621 TACGACCCCGAGAGAGAGCTGTGCGCGAGATCTGAGAACGAGGCGCACAGCATGTGAGC 1680
Qy 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACCCCAAGATCCG 1740
Db 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAGAACCGGCAAGTACCCCAAGATCCG 1740
Qy 1741 ACCGCGCACACCAAGAGCTGAAAGAGCTGTGACCGAGCGGTGTGAGAAATCCCATGTAG 1800
Db 1741 ACCGCGCACACCAAGAGCTGAAAGAGCTGTGACCGAGCGGTGTGAGAAATCCCATGTAG 1800
Qy 1801 AAGCATGTGATCTGGGGCAAGACCCCAAGTCCCGCTGCCATCCAGAGAGAGACTGTG 1860
Db 1801 AAGCATGTGATCTGGGGCAAGACCCCAAGTCCCGCTGCCATCCAGAGAGAGACTGTG 1860
Qy 1861 GAGACCTGTGAGCCGACTACTGTGAGGACCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920
Db 1861 GAGACCTGTGAGCCGACTACTGTGAGGACCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920
Qy 1921 ACCCCCCCTGTGTGAAGCTGTGTATCAGCTGTGAGAGAGAGCCCATCATCTGCGCGCAG 1980
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTATCAGCTGTGAGAGAGAGCCCATCATCTGCGCGCAG 1980
Qy 1981 ACTTTTAAAGTGTGAGCGCGCGCGCGCAACCGGAGAGACCAAGATCGGCAAGGCTGTGCTG 2040
Db 1981 ACTTTTAAAGTGTGAGCGCGCGCGCGCGCAACCGGAGAGACCAAGATCGGCAAGGCTGTGCTG 2040
Qy 2041 ACCGACCGGGGCGCGGCAAGAGCTGTGAGCTGTGACGAGACGAGGAGTGAACATGTGTACGAC 2100
Db 2041 ACCGACCGGGGCGCGGCAAGAGCTGTGAGCTGTGACGAGACGAGGAGTGAACATGTGTACGAC 2100
Qy 2101 CTGCAAGGCGCATTCAGCTGTGCGCTGTGAGAGACGCGGACGAGGTGAACATCGTGAACGAC 2160
Db 2101 CTGCAAGGCGCATTCAGCTGTGCGCTGTGAGAGACGCGGACGAGGTGAACATCGTGAACGAC 2160
Qy 2161 AAGCAATAGCGCTTGGGCGATCATCCAGGCGCCAGCCCGACAAAGGCGAGGCGAGCTGTG 2220
Db 2161 AAGCAATAGCGCTTGGGCGATCATCCAGGCGCCAGCCCGACAAAGGCGAGGCGAGCTGTG 2220

```

```

Qy 2221 AACCATGATCTGAGCAGCTGATCAAGAGAGAGGCTACTGTAGCTGTGTCGCGC 2280
Db 2221 AACCATGATCTGAGCAGCTGATCAAGAGAGAGGCTACTGTAGCTGTGTCGCGC 2280
Qy 2281 CACAGGCGCATTCGCGCGCAAGAGAGATTCAGCAAGCTGTGTGAGCAAGGCGCATCCGCAAG 2340
Db 2281 CACAGGCGCATTCGCGCGCAAGAGAGATTCAGCAAGCTGTGTGAGCAAGGCGCATCCGCAAG 2340
Qy 2341 GTGCTGTTCGTGACGCGCATGATGCGCGCATGTGATCTACCACTACATGACGACCTG 2400
Db 2341 GTGCTGTTCGTGACGCGCATGATGCGCGCATGTGATCTACCACTACATGACGACCTG 2400
Qy 2401 TACGTGGGCAAGCGCGCGCTTGTGATTTAAAGCTTCCGCGGCTTACGACCCGCT 2457
Db 2401 TACGTGGGCAAGCGCGCGCTTGTGATTTAAAGCTTCCGCGGCTTACGACCCGCT 2457

RESULT 10
ADCI3266
ID ADCI3266 standard; DNA; 2457 BP.
XX
XX
AC ADCI3266;
XX
XX
DT 18-DEC-2003 (first entry)
XX
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
OS Human immunodeficiency virus.
XX
PN WO200304620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SM, Lian Y, Engelbrecht S, Van Renenburg EJ,
XX
XX WPI; 2003-221593/21.
XX
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunisation, or generating
PT packaging cell lines.
XX
XX
PS Disclosure; Fig 42; 301pp; English.
XX
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
XX
SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

Query Match 98.7%; Score 2430.2; DB 10; Length 2457;
Best Local Similarity 99.6%; Pred. No. 5.2e-293;

```

Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 7 GCCACCATGCGCGAGGCATGAGCCAGGCGACCAAGCGCCAAATCTGTGATGAGCGCGAC 66
Db 1 GCCACCATGCGCGAGGCATGAGCCAGGCGACCAAGCGCCAAATCTGTGATGAGCGCGAC 60
QY 67 AACTTCAGAGGCGCCCAAGCGCATCATCATGCTTCAACTGCGGCAAGAGGCGCCACATC 126
Db 61 AACTTCAGAGGCGCCCAAGCGCATCATCATGCTTCAACTGCGGCAAGAGGCGCCACATC 120
QY 127 GCCCGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
Db 121 GCCCGCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 187 CAGATGAGAGGATGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 246
Db 181 CAGATGAGAGGATGCAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 247 CAGGCGCAAGGCGCGCGAGTTCCCGAGCGAGCAAGAACCGCGCGCAAGCGCGCGCGCG 306
Db 241 CAGGCGCAAGGCGCGCGAGTTCCCGAGCGAGCAAGAACCGCGCGCAAGCGCGCGCGCG 300
QY 307 GAGCTGCAAGGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCG 366
Db 301 GAGCTGCAAGGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCG 360
QY 367 CTGAACTTCCCGCAGATCACTCGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 361 CTGAACTTCCCGCAGATCACTCGTGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 427 CAGATCAAGAGAGCGCGCTGTGACACCGCGCGCGAGCAACCGGTGTGAGAGATGAGC 486
Db 421 CAGATCAAGAGAGCGCGCTGTGACACCGCGCGCGAGCAACCGGTGTGAGAGATGAGC 480
QY 487 CTGCGCGCGCAATGTGAAGCCCAAGATGATTCGCGCGCATTCGCGCTTCACTAAGGTGCG 546
Db 481 CTGCGCGCGCAATGTGAAGCCCAAGATGATTCGCGCGCATTCGCGCTTCACTAAGGTGCG 540
QY 547 CAGTACGACACAGATCTGTGATCGAGATCTGCGGCAAGAGCCATCGGCGCGGTGCTGATC 606
Db 541 CAGTACGACACAGATCTGTGATCGAGATCTGCGGCAAGAGCCATCGGCGCGGTGCTGATC 600
QY 607 GCGCGCGCGCGCGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGCGCTGACCCCTG 666
Db 601 GCGCGCGCGCGCGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGCGCTGACCCCTG 660
QY 667 AACTTCCCATCAAGCGCGCATTCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGACCG 726
Db 661 AACTTCCCATCAAGCGCGCATTCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGACCG 720
QY 727 CCGAAGGTGAAGAGTGGCGCGTGAACCGAGAGAGATGCAAGCGCGTGAACCGCGCATTCG 786
Db 721 CCGAAGGTGAAGAGTGGCGCGTGAACCGAGAGAGATGCAAGCGCGTGAACCGCGCATTCG 780
QY 787 GAGGAGATGAGAGAGAGGCGCAAGATCAACCAAGATTCGCGCGCGAGAACCCCTTCAACAAC 846
Db 781 GAGGAGATGAGAGAGAGGCGCAAGATCAACCAAGATTCGCGCGCGAGAACCCCTTCAACAAC 840
QY 847 CCGGTGTTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
Db 841 CCGGTGTTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 907 GAGCTGAACAAGCGCGCGAGGACTTCTGTGAGGTGCAAGCTGAGGACTTCCCGACCGCGCG 966
Db 901 GAGCTGAACAAGCGCGCGAGGACTTCTGTGAGGTGCAAGCTGAGGACTTCCCGACCGCGCG 960
QY 967 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
Db 961 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1027 CCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
Db 1021 CCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080

QY 1087 ACCCGCGCATTCGCTACCAAGTACCAAGTGTGCTGCCAGAGGCTGGAAGGCGAGCCCGAC 1146
Db 1081 ACCCGCGCATTCGCTACCAAGTACCAAGTGTGCTGCCAGAGGCTGGAAGGCGAGCCCGAC 1140
QY 1147 ATCTTCAGAGAGAGATGCAAGATCTGTGAAGCTTTCGCGCGCGCAAGCCCGAGATC 1206
Db 1141 ATCTTCAGAGAGAGATGCAAGATCTGTGAAGCTTTCGCGCGCGCAAGCCCGAGATC 1200
QY 1207 GTGATCTACCA-----GCGCGCGCTGTACGTGGGCAAGGAGAGATGGGCGAGAC 1260
Db 1201 GTGATCTACCAAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 CCGCGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 CCGCGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Db 1321 AAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 TGGACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 TGGACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 AAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1441 AAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 1501 CTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 1561 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db 1681 TACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 AGCATGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1801 AGCATGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 GAGACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1861 GAGACCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 ACCCGCGCGCGTGTGAAGCTGTGTGAACAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 ACCCGCGCGCGTGTGAAGCTGTGTGAACAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 ACCCTTCTAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db 1981 ACCCTTCTAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 ACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2041 ACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160


```
Db 961 GGCCTGAAGAAAGAGAGCGTGAACCGTGTGAGCTGGGCGAGCGCTACTTCAAGCGG 1020
Qy 1027 CCCCTGAGACGAGACTTCCTCGCAAGTACACCGGCTTCAACATCCCGAGCATCAACAGG 1086
Db 1021 CCCCTGAGACGAGACTTCCTCGCAAGTACACCGGCTTCAACATCCCGAGCATCAACAGG 1080
Qy 1087 ACCCGCGGATCCGCTACCAAGTACCAAGTGTGCCCAAGGGCTGGAAGGGCAGCCAGC 1146
Db 1081 ACCCGCGGATCCGCTACCAAGTACCAAGTGTGCCCAAGGGCTGGAAGGGCAGCCAGC 1140
Qy 1147 ATCTTCAGAGCAGATGACCAAGATCTGAGAGCCCTTCGCGCCGCAACCCGAGATC 1206
Db 1141 ATCTTCAGAGCAGATGACCAAGATCTGAGAGCCCTTCGCGCCGCAACCCGAGATC 1200
Qy 1207 GTGATCTACGAGCCCGCTGTACAGTGGGCAAGACCTGGAAATGGGCGCAACCGCGC 1266
Db 1201 GTGATCTACGAGCCCGCTGTACAGTGGGCAAGACCTGGAAATGGGCGCAACCGCGC 1260
Qy 1267 AAGATCGAGAGCTGGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGACAAGAG 1326
Db 1261 AAGATCGAGAGCTGGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGACAAGAG 1320
Qy 1327 CACGAGAAAGAGAGCCCGCTTCTGTGATGGCTACAGAGCTGCAACCCGACAAGTGAAC 1386
Db 1321 CACGAGAAAGAGAGCCCGCTTCTGTGATGGCTACAGAGCTGCAACCCGACAAGTGAAC 1374
Qy 1387 GTGCAAGCCCATGTGAGCTGCCGCAAGAGAGAGAGCTGACCTGTGAAGACATCCAGAACTG 1446
Db 1375 GTGCAAGCCCATGTGAGCTGCCGCAAGAGAGAGAGCTGACCTGTGAAGACATCCAGAACTG 1434
Qy 1447 GTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCAATGAAGTGGCGAGCTGTC 1506
Db 1435 GTGGGCAAGCTGAACCTGGGCGAGCCAGATCTACCCCGGCAATGAAGTGGCGAGCTGTC 1494
Qy 1507 AAGCTGTGCGGCGCGCAAGGCGCTGACCGACATGCTGCGCCCTGACCGAGAGGCGGAG 1566
Db 1495 AAGCTGTGCGGCGCGCAAGGCGCTGACCGACATGCTGCGCCCTGACCGAGAGGCGGAG 1554
Qy 1567 CTGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGCTGCAAGGCTGTACTACAG 1626
Db 1555 CTGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGCTGCAAGGCTGTACTACAG 1614
Qy 1627 CCCAGCAAGAGACTGGTGGCGGAGATCCAGAAAGAGGCGCAAGACGATGAGACTTACAG 1686
Db 1615 CCCAGCAAGAGACTGGTGGCGGAGATCCAGAAAGAGGCGCAAGACGATGAGACTTACAG 1674
Qy 1687 ATCTACGAGAGCCCTTCAAGAACTGAAAGCCGGCAAGTACGCCAAGATGGCGACCGCC 1746
Db 1675 ATCTACGAGAGCCCTTCAAGAACTGAAAGCCGGCAAGTACGCCAAGATGGCGACCGCC 1734
Qy 1747 CACACCAAGAGAGTGAAGAGAGCTGACCGAGCGCTGCAAGATGCGCATGAGAGCATC 1806
Db 1735 CACACCAAGAGAGTGAAGAGAGCTGACCGAGCGCTGCAAGATGCGCATGAGAGCATC 1794
Qy 1807 GTGATCTGGGCGAAGACCCCGAAGTTCGGCTGCGCATCCAGAAAGAGACTGGGAGATC 1866
Db 1795 GTGATCTGGGCGAAGACCCCGAAGTTCGGCTGCGCATCCAGAAAGAGACTGGGAGATC 1854
Qy 1867 TGTGTGACCGACTACTGTGAGAGGCACTTGATCCCGAGTGGAGATTCTGTAAACCCCC 1926
Db 1855 TGTGTGACCGACTACTGTGAGAGGCACTTGATCCCGAGTGGAGATTCTGTAAACCCCC 1914
Qy 1927 CCCCTGTGAAGTGTGTGTACCAAGCTGGAAGAAAGAGCCCATCATCGGCGCCGAGACTTTC 1986
Db 1915 CCCCTGTGAAGTGTGTGTGTACCAAGCTGGAAGAAAGAGCCCATCATCGGCGCCGAGACTTTC 1974
Qy 1987 TACGTGGAAGGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGAGCCGAC 2046
Db 1975 TACGTGGAAGGGCGCGCAACCGCGAGACCAAGATCGGCAAGCGCGCTACGAGCCGAC 2034
Qy 2047 CGGGGCGCGCAAGATCTGTAGCGCTGACCGGAGACCAACCAAGAGCCGAGCTGACG 2106
```

```
Db 2035 CGGGGCGCGCAAGATCTGTAGCGCTGACCGAGACCAACCAAGAGACGAGCTGACG 2094
Qy 2107 GGCATCGAGCTGGGCGCTGCAAGACAGCGCGGAGGTGAACATGTGTACCGACAGCCAG 2166
Db 2095 GGCATCGAGCTGGGCGCTGCAAGACAGCGCGGAGGTGAACATGTGTACCGACAGCCAG 2154
Qy 2167 TACGCGCTGGGATCATCAAGGCCAGCCGCAAGAGAGAGAGAGAGCTGTGAAACAG 2226
Db 2155 TACGCGCTGGGATCATCAAGGCCAGCCGCAAGAGAGAGAGAGAGCTGTGAAACAG 2214
Qy 2227 ATCATGAGCAGCTGATCAAGAGAGAAAGTGTACTGTAGCTGGTGGCCCGCAAG 2286
Db 2215 ATCATGAGCAGCTGATCAAGAGAGAAAGTGTACTGTAGCTGGTGGCCCGCAAG 2274
Qy 2287 GGCATCGGCGCAAGACAGATGCAAGAGCTGTGAGCAAGGCGCATCCGAAAGTGTCTG 2346
Db 2275 GGCATCGGCGCAAGACAGATGCAAGAGCTGTGAGCAAGGCGCATCCGCAAGGTGTCTG 2334
Qy 2347 TTCTGTGAGCGGATGATGCGGCGCATGCTGTACTACAGTACATGAGACCTGTACCTG 2406
Db 2335 TTCTGTGAGCGGATGATGCGGCGCATGCTGTACTACAGTACATGAGACCTGTACCTG 2394
Qy 2407 GGCAGCGGCGGCGCTTAGAGTCAATTAAGCTTCCCGGCGCTAGACCGGT 2457
Db 2395 GGCAGCGGCGGCGCTTAGAGTCAATTAAGCTTCCCGGCGCTAGACCGGT 2445

RESULT 12
ADCI3264
ID ADCI3264 standard; DNA; 2445 BP.
XX
XX
AC ADCI3264;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YMMW_C SEQ ID NO 43.
XX
KM expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prox;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; da.
XX
OS Human immunodeficiency virus.
XX
PN MO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR ) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SM, Llan Y, Engelbrecht S, Van Rensburg BJ;
XX
XX WPI; 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX Prox, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 40; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prox, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
```

CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 10; Length 2445;
Best Local Similarity 99.6%; Pred. No. 4.5e-292;
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

```
QY 7 GGCACCATGCGCGAGCCCATGAGCCAGGCGCACCGAGGCCAACAATCCTGTATGACGGCAGC 66
DB 1 GGCACCATGCGCGAGCCCATGAGCCAGGCGCACCGAGGCCAACAATCCTGTATGACGGCAGC 60
QY 67 AACTTCAGAGGGCCCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCGCAATC 126
DB 61 AACTTCAGAGGGCCCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCGCAATC 120
QY 127 GCCCGCACTGCCGCGCCCGCCCGCAAGAGGGCTGCTGAAAGTGGCGCAAGAGGGCGCAC 186
DB 121 GCCCGCACTGCCGCGCCCGCCCGCAAGAGGGCTGCTGAAAGTGGCGCAAGAGGGCGCAC 180
QY 187 CAGATGAAGACTGCAACCGAGGCGCAGGCGCACTTCTTCGCGAGGACCTGCGCTTCC 246
DB 181 CAGATGAAGACTGCAACCGAGGCGCAGGCGCACTTCTTCGCGAGGACCTGCGCTTCC 240
QY 247 CAGGCGAAGGCGCGCGAGTTCCCGACGAGCAAGACCGGCGCAACAGCCCAACAGCCGC 306
DB 241 CAGGCGAAGGCGCGCGAGTTCCCGACGAGCAAGACCGGCGCAACAGCCCAACAGCCGC 300
QY 307 GAGCTGCAAGTGGCGGCGCAACAACCCCGCAGCGGAGGCGCGCGCGCGCGCGCGCGCG 366
DB 301 GAGCTGCAAGTGGCGGCGCAACAACCCCGCAGCGGAGGCGCGCGCGCGCGCGCGCGCG 360
QY 367 CTGAAATTTCCCGCAAGTCAACCTGTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
DB 361 CTGAAATTTCCCGCAAGTCAACCTGTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 427 CAGATCAAGAGGCGCTGTGAGCAACCGGCGCGAGCGACACACGCTGTGAGAGAGTAGC 486
DB 421 CAGATCAAGAGGCGCTGTGAGCAACCGGCGCGAGCGACACACGCTGTGAGAGAGTAGC 480
QY 487 CTGCGCGGCAAGTGAAGCGCAAGATGATGCGCGGCAATCGGCGGCTTCAACAGGTGCG 546
DB 481 CTGCGCGGCAAGTGAAGCGCAAGATGATGCGCGGCAATCGGCGGCTTCAACAGGTGCG 540
QY 547 CAGTACGACCAAGTCTGTATCGAGATCTGCGGCAAGAGGCGCATCGGCACTGTGTGATC 600
DB 541 CAGTACGACCAAGTCTGTATCGAGATCTGCGGCAAGAGGCGCATCGGCACTGTGTGATC 600
QY 607 GGGCCCAACCCCGGTGAACATCATCGGCGGCAACATGCTGAACCCGAGCTGAGCTGACCTG 666
DB 601 GGGCCCAACCCCGGTGAACATCATCGGCGGCAACATGCTGAACCCGAGCTGAGCTGACCTG 660
QY 667 AACTTCCCATCAGCCCATCAGACCGTGCCTGTGAAGCTGAAGCCCGCATGAAGCGC 726
DB 661 AACTTCCCATCAGCCCATCAGACCGTGCCTGTGAAGCTGAAGCCCGCATGAAGCGC 720
QY 727 CCCAAGGTGAAGAGTGGCCCTTGAACGAGAGAAAGATCAAGAGCCCTTGAACCGCATTTGC 786
DB 721 CCCAAGGTGAAGAGTGGCCCTTGAACGAGAGAAAGATCAAGAGCCCTTGAACCGCATTTGC 780
QY 787 GAGAGATGAGAGAGAGGCGCAAGATCAACCAAGTGGCGCGCGAGAACCCCTTCAACACC 846
DB 781 GAGAGATGAGAGAGAGGCGCAAGATCAACCAAGTGGCGCGCGAGAACCCCTTCAACACC 840
QY 847 CCGGTGTTCCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGTGTGAATTCGCG 906
DB 841 CCGGTGTTCCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGTGTGAATTCGCG 900
QY 907 GAGGTGAACAAAGCGCACCCAGAACTTTGGAGGTGAGCTGGGCAATCCCCCAACCCCGCC 966
```

```
DB 901 GAGCTGAACAAGCGCACCCAGGACTTTCGGAGAGTGGACGTGGGACATCCCCACCCCGCC 960
QY 967 GGCCTGAAGAGAGAGAGGCGTGAACCTGTCTGAGACGTGGGCGAGCGCTTACCTTCAAGCTG 1026
DB 961 GGCCTGAAGAGAGAGAGGCGTGAACCTGTCTGAGACGTGGGCGAGCGCTTACCTTCAAGCTG 1020
QY 1027 CCCCTGAGACGAGGACTTCCGCAAGTACACGCGCTTCAACCATCCCGAGCATCAACAGAG 1086
DB 1021 CCCCTGAGAGAGGACTTCCGCAAGTACACGCGCTTCAACCATCCCGAGCATCAACAGAG 1080
QY 1087 ACCCGCGCATCCGCTACCAAGTACCAAGTGTCTGCCCAAGGCGTGAAGAGGCGAGCCCGAG 1146
DB 1081 ACCCGCGCATCCGCTACCAAGTACCAAGTGTCTGCCCAAGGCGTGAAGAGGCGAGCCCGAG 1140
QY 1147 ATCTTCAGAGCAATGACCAAGATCTGTAGGCTTTCGCGCGCGCAACCCGAGATC 1206
DB 1141 ATCTTCAGAGCAATGACCAAGATCTGTAGGCTTTCGCGCGCGCAACCCGAGATC 1200
QY 1207 GTGATCTACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266
DB 1201 GTGATCTACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1267 AAGATCGAGAGCTGCGCAAGCACTGTGCGCTGGGCGTTCAACAACCCCGCAAGAGAG 1326
DB 1261 AAGATCGAGAGCTGCGCAAGCACTGTGCGCTGGGCGTTCAACAACCCCGCAAGAGAG 1320
QY 1327 CACCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
DB 1321 CACCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1374
QY 1387 GTGCGAGCCCATGAGAGTGGCGCGAGAGAGAGAGTGGACCGTGAACGATCCCAAGAGCTG 1446
DB 1375 GTGCGAGCCCATGAGAGTGGCGCGAGAGAGAGAGTGGACCGTGAACGATCCCAAGAGCTG 1434
QY 1447 GTGGGCAAGCTGAACCTGGGCGAGCGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGC 1506
DB 1435 GTGGGCAAGCTGAACCTGGGCGAGCGCAGATCTACCCCGGATCAAGGTGGCGCAGCTGTGC 1494
QY 1507 AAGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1566
DB 1495 AAGCTGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1554
QY 1567 CTGAGAGTGGCGGAGAACCGCGGAGATCTGTGCGGAGACCCGTGACCGCGGTGACTAGAC 1626
DB 1555 CTGAGAGTGGCGGAGAACCGCGGAGATCTGTGCGGAGACCCGTGACCGCGGTGACTAGAC 1614
QY 1627 CCCAGCAAGACCTGTGTGCGCGAGATCCAGAAAGCAGAGGCGACGACAGTGAAGCTTACAG 1686
DB 1615 CCCAGCAAGACCTGTGTGCGCGAGATCCAGAAAGCAGAGGCGACGACAGTGAAGCTTACAG 1674
QY 1687 ATCTACCAAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGACCGCC 1746
DB 1675 ATCTACCAAGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGACCGCC 1734
QY 1747 CACACCAAGAGCTGAAGAGCTGACCGAGGCGGTGCAAGAAATTCGCAATGAGAGATC 1806
DB 1735 CACACCAAGAGCTGAAGAGCTGACCGAGGCGGTGCAAGAAATTCGCAATGAGAGATC 1794
QY 1807 GTGATCTGGGGCAAGACCCCGAAGTTCCGCGCTGCGCATCCAGAGAGAGACTTGGAGAGC 1866
DB 1795 GTGATCTGGGGCAAGACCCCGAAGTTCCGCGCTGCGCATCCAGAGAGAGACTTGGAGAGC 1854
QY 1867 TGGTGAACGACCTTACGAGGCGACCTGAGATCCCGAGTGGAGTTCTGTAACACCCCG 1926
DB 1855 TGGTGAACGACCTTACGAGGCGACCTGAGATCCCGAGTGGAGTTCTGTAACACCCCG 1914
QY 1927 CCCCTGTGAAGCTGTGTGATCCAGCTGAGAGAGAGCCCATCATCGCGCGCGGACCTTTC 1986
DB 1915 CCCCTGTGAAGCTGTGTGATCCAGCTGAGAGAGAGCCCATCATCGCGCGCGGACCTTTC 1974
QY 1987 TACGTGAACGAGCGCGCGCAACCGCGAGACCAAGATGCGGCAAGGCGCGGCTTACGTGACGAC 2046
DB 1975 TACGTGAACGAGCGCGCGCGCAACCGCGAGACCAAGATGCGGCAAGGCGCGGCTTACGTGACGAC 2034
```


OY	913	CAAAGCCACCAAGACTTCTCGGAGAGTGACGCTGGGCAATCCCAACCCGCGGGCTTGA	973
Db	2387	ACAAGCCGACCACAGACTTCTGGSAGAGGTGACGCTGGGCAATCCCAACCCGCGGGCTTGA	2446
OY	974	AGAAAGAAAGAGGGTGAACCTGCTGAGACGTGGGCGAAGGCTACTTTCAGGGCTGCCCTG	1033
Db	2447	AGAAAGAAAGAGGGTGAACCTGCTGAGACGTGGGCGAAGGCTACTTTCAGGGCTGCCCTG	2506
OY	1034	ACGAGGACTTTCGGCAAGTACACCGCTTTCACCATCCCAAGATCAACAAGAACCCCG	1093
Db	2507	ACGAGGACTTTCGGCAAGTACACCGCTTTCACCATCCCAAGATCAACAAGAACCCCG	2566
OY	1094	GCATCCGCTACAGTACAAAGTGTGCGCCCGGAGGCTGAAAGGGGAGGCCCAAGACTTCTC	1153
Db	2567	GCATCCGCTACAGTACAAAGTGTGCGCCCGGAGGCTGAAAGGGGAGGCCCAAGACTTCTC	2626
OY	1154	AGAGCAGCATGACCAAGATCCTGAGCCCTTTCGGGCGCCGACACCCGAGATGTGATCT	1213
Db	2627	AGAGCAGCATGACCAAGATCCTGAGCCCTTTCGGGCGCCGACACCCGAGATGTGATCT	2686
OY	1214	ACCAAGCCCCCTCTTACGTGGGCGAGGCACTTGAGATTCGGCAGACCGGGCCAAAGTTCG	1273
Db	2687	ACCAAGCCCCCTCTTACGTGGGCGAGGCACTTGAGATTCGGCAGACCGGGCCAAAGTTCG	2746
OY	1274	AGGAGCTGCGCAACACCTGCTGCGCTGGGGCTTCAACAACCCCGCAAGAAAGCACAGA	1333
Db	2747	AGGAGCTGCGCAACACCTGCTGCGCTGGGGCTTCAACAACCCCGCAAGAAAGCACAGA	2806
OY	1334	AGGAGCCCCCTTCTCTGTGATGGGCTTACGAGCTGACACCCCGACAGTGAACGTCGACG	1393
Db	2807	AGGAGCCCCCTTCTCTGTGATGGGCTTACGAGCTGACACCCCGACAGTGAACGTCGACG	2860
OY	1394	CCATCGAGCTGCCGAGAAAGGAGGCTGGAACCGTGAACGACATCCAGAGCTGGTGGACA	1453
Db	2861	CCATCGAGCTGCCGAGAAAGGAGGCTGGAACCGTGAACGACATCCAGAGCTGGTGGACA	2920
OY	1454	AGCTGAATCTGGGCGACGCCAGATCTTACCCCGGCAATCAAGTGGCGCAGCTGTGCAAGCTGC	1513
Db	2921	AGCTGAATCTGGGCGACGCCAGATCTTACCCCGGCAATCAAGTGGCGCAGCTGTGCAAGCTGC	2980
OY	1514	TGCGCGGGCCCAAGGCCCTTGAACGACATCTGTCGCCCTGACCGAAGAGAGCGAGCTGAGC	1573
Db	2981	TGCGCGGGCCCAAGGCCCTTGAACGACATCTGTCGCCCTTGAACCGAAGAGAGCGAGCTGAGC	3040
OY	1574	TGGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGGTGACAGCGGTGTACTTACGACCCACGA	1633
Db	3041	TGGCGGAGAACCGCGAGATCTCTGCGCGAGCCCGGTGACAGCGGTGTACTTACGACCCACGA	3100
OY	1634	AGGACCTGTGTGACCGAGATTCAGAAAGCAGGGCCACGACCACTGACCTTACGATCTTACC	1693
Db	3101	AGGACCTGTGTGACCGAGATTCAGAAAGCAGGGCCACGACCACTGACCTTACGATCTTACC	3160
OY	1694	AGGAGGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAABAATGGGACACGGCCACACGA	1753
Db	3161	AGGAGGCCCTTCAAGAACCTGAAAGACCGGCAAGTACGCCAABAATGGGACACGGCCACACGA	3220
OY	1754	ACGACGTGAGCAGCTGACCGAGGCGGTGCAGAAAGATTCGCCATGAGAGCATCTGTGATCT	1813
Db	3221	ACGACGTGAGCAGCTGACCGAGGCGGTGCAGAAAGATTCGCCATGAGAGCATCTGTGATCT	3280
OY	1814	GGGGCAAGACCCCAAGTTTCGGCTGTGCCATTCAGAAAGAGACCTGGGAGACCTGTGTGA	1873
Db	3281	GGGGCAAGACCCCAAGTTTCGGCTGTGCCATTCAGAAAGAGACCTGGGAGACCTGTGTGA	3340
OY	1874	CCGACTTACGAGGCGACACTTGTATCCCGGAGTGGGAGTTGTGTAACAACCCCCCGCTGG	1933
Db	3341	CCGACTTACGAGGCGACACTTGTATCCCGGAGTGGGAGTTGTGTAACAACCCCCCGCTGG	3400
OY	1934	TGAAGCTGTGGTACAGCTGCGAGAAAGACCCATCATCTGGCGCGGAGACTTCTTACTGTGG	1993
Db	3401	TGAAGCTGTGGTACAGCTGCGAGAAAGACCCATCATCTGGCGCGGAGACTTCTTACTGTGG	3460

QY	1994	ACGGCGCCGCGCAACCGGGAACCAAGTCGCAAGCGCGGCTACGTGACCGACCGGGGCG	2053
Db	3461	ACGGCGCGCGCAACCGGGAACCAAGTCGCAAGCGCGGCTACGTGACCGACCGGGGCG	3520
QY	2054	GGCAGAAGATCGTAGGCTTGACCGAGACCAACCAAGAGACCGAGCTGACGACATCC	2113
Db	3521	GGCAGAAGATCGTAGGCTTGACCGAGACCAACCAAGAGACCGAGCTGACGAGCATCC	3580
QY	2114	AGCTGGCCCTGCAAGACACGCGGACGCGAGGTGAACATCGTAGCCGACAGCCATGACGCC	2173
Db	3581	AGCTGGCCCTGCAAGACACGCGGACGCGAGGTGAACATCGTAGCCGACAGCCATGACGCC	3640
QY	2214	TGGGCATCATCAGAGCCGAGCCGACCAAGAGCGGAGCGAGCTGATGAACCAATCATCTCG	2233
Db	3641	TGGGCATCATCAGAGCCGAGCCGACCAAGAGCGGAGCGAGCTGATGAACCAATCATCTCG	3700
QY	2234	AGCAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGGTGCCCGCCACAAAGGGCATCG	2293
Db	3701	AGCAGCTGATCAAGAGAGAGAGGTGTACTGAGCTGGGTGCCCGCCACAAAGGGCATCG	3760
QY	2254	GCGGCAACGAGCAATGAGCAAGCTGTGTAGCAAGGGCATCCGCAAGGTGTCTCTCG	2353
Db	3761	GCGGCAACGAGCAATGAGCAAGCTGTGTAGCAAGGGCATCCGCAAGGTGTCTCTCG	3820
QY	2354	ACGGCATTCGATGCGGCGATCGTGATTCACAGATTCATGAGACGACCTTACGTGGGACGCG	2413
Db	3821	ACGGCATTCGATGCGGCGATCGTGATTCACAGATTCATGAGACGACCTTACGTGGGACGCG	3880
QY	2414	GCGGCGCTAGGATCGATTTAAAGCTTCCCGGGGCTAGACCCGGT	2457
Db	3881	GCGGCGCTAGGATCGATTTAAAGCTTCCCGGGGCTAGACCCGGT	3924

XX	RESULT 14
ADCI3231	ID ADCI3231 standard; DNA; 3930 BP.
AC	ADCI3231;
DT	18-DEC-2003 (first entry)
DE	DNA of HIV construct GagCompIPolmutA/C_C SEQ ID NO 10.
KM	expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; ProC; Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
OS	Human immunodeficiency virus.
PN	WO2003004620-A2.
PD	16-JAN-2003.
PF	05-JUL-2002; 2002MO-USO21420.
PR	05-JUN-2001; 2001US-0303192P. 31-AUG-2001; 2001US-0316860P.
PP	16-JAN-2002; 2002US-0349871P.
PA	(CHTR) CHIRON CORP. (UYST-) UNIV STELLENBOSCH.
Zur Megede J,	Barnett SW, Llan Y, Engelbrecht S, Van Rensburg EJ, WPI; 2003-221593/21.
New expression cassette comprising a polynucleotide sequence encoding a polyprotein including an HIV gag, env, int, nef, p15RNaseH, pol, tat, proC, or Rev polyprotein, useful for immunization, or generating packaging cell lines.	
The invention relates to a novel expression cassette comprising a	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8697.67 Seconds
(without alignments) 13249.138 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463

Sequence: 1 gtcgacgcaccacatgscgca.....gggctagcaccgctgaattc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*
10: gb_esc10:*
11: gb_esc11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136.4	5.5	330	10	CL293849 02S0349-0
2	98.4	4.0	2886	10	CL967755 OaIFCC015
3	90.2	3.7	1781	8	CK099623 RRCm0260
4	86.8	3.5	2598	4	AY103647 Zea maye
5	86.6	3.5	1398	10	CL247380 OaIFCC006
6	86	3.5	743	10	CZ247380 AIAA-af3
7	85.4	3.5	951	3	BM321451 rockefell
8	85.4	3.5	2031	10	CL974989 OaIFCC042
9	84	3.4	1725	10	CL978463 OaIFCC031
10	80.8	3.3	1060	10	CW922203 BDCAR297R
11	80.8	3.3	3069	10	CL973391 OaIFCC025
12	80.6	3.3	892	10	CK159167 FGA504056
13	80.6	3.3	892	10	CZ216254 AIAA-af2
14	80.6	3.3	1941	10	CL971508 OaIFCC021
15	80.2	3.3	1132	3	BM320864 rockefell
16	79.6	3.2	1165	3	BM320900 rockefell
17	79.6	3.2	11691	10	CL962901 OaIFCC008
18	79	3.2	867	3	BM321430 OaIFCC017
19	78.6	3.2	2682	10	CL969033 OaIFCC017
20	78.4	3.2	873	7	CO443531 MGCCL1006
21	78.2	3.2	2853	10	CL974387 OaIFCC025
22	78	3.2	1485	10	CL970981 OaIFCC020

23	78	3.2	2313	10	CL982362	CL982362 OaIFSC047
24	78	3.2	2454	10	CL975440	CL975440 OaIFCC027
25	77.4	3.1	1509	10	CL959255	CL959255 OaIFCC002
26	77.4	3.1	2559	10	CL982027	CL982027 OaIFSC046
27	76.6	3.1	1550	3	BM321022	BM321022 rockefell
28	76.4	3.1	2028	10	CL979437	CL979437 OaIFCC033
29	76.2	3.1	1401	10	CL962721	CL962721 OaIFCC038
30	76.2	3.1	1680	10	CL982770	CL982770 OaIFSC049
31	76	3.1	2793	10	CL948321	CL948321 OaIFSC004
32	75	3.0	862	3	BM321023	BM321023 rockefell
33	75	3.0	1734	10	CL948553	CL948553 OaIFSC005
34	74.6	3.0	1410	10	CL947314	CL947314 OaIFSC004
35	74.6	3.0	1989	10	CL980973	CL980973 OaIFSC035
36	74.4	3.0	914	9	BZ568300	BZ568300 pac82-164
37	74	3.0	1071	10	CW922105	CW922105 BDCB195TF
38	74	3.0	1290	10	CL972679	CL972679 OaIFCC023
39	74	3.0	1386	11	DQ045165	DQ045165 Homo sapi
40	74	3.0	2072	4	CR603312	CR603312 full-length
41	73.2	3.0	853	3	BM321393	BM321393 rockefell
42	73	3.0	2151	10	CL972100	CL972100 OaIFCC041
43	72.8	3.0	566	3	BM587428	BM587428 170006873
44	72.6	2.9	788	6	CB643171	CB643171 OSUNB03L
45	72.4	2.9	753	9	CC675888	CC675888 OGMCO51TH

ALIGNMENTS

RESULT 1
LOCUS CL293849/c 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MutRIL Library Zea maye genomic clone
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea maye
ORGANISM Zea maye
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 330)
REFERENCE Lathaw, S., Tan, B.-C., Settles, A.M. and McCarty, D.R.
AUTHORS Sequence tagged transposon insertions from the UniformMu maize population
TITLE Unpublished (2003)
JOURNAL Contact: Donald R. McCarty
COMMENT Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmcufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
02S0349-08, Primer set: A
Class: transposon insertion site.
Location/Qualifiers
1..330
/organism="Zea maye"
/mol_type="genomic DNA"
/strain="w22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sephadex 400 spin columns and cloned into the TOPO PCR4 vector."

FEATURES

source

ORIGIN

Query Match	5.5%	Score 136.4	DB 10	Length 330
Best Local Similarity	66.4%	Pred. No. 2.7e-14		
Matches 211	Conservative	0	Mismatches 106	Indels 1
QY	421	GGCGGCGCAGATCAAGAGGCGCTCTGTGACACCGGCGCGACGACCGCTGTGAGGAG	480	
Db	329	GGGGGGCAGCTGGAGAGAGCTTATTAGATACAGAGCGAGATGATACGATATTAGAGAA	270	
QY	481	ATGAGCCTGCGCCGGCAGTGGAGAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG	540	
Db	269	ATGACTTTTGAACAGAGAAAGATGAAACCAAAAATGATAGGGGGAATTGGAGTTTTATCAA	210	
QY	541	GTGGGCGCAGTACAGACCAATCTCTGATCGAGATCTGCGCAGAAAGGCATCGGCACCGTG	600	
Db	209	GTAAACACGTATGATGATCAGGTACCCATAGAAATCTGTGGGCATTAAGCTATAGTACGGTA	150	
QY	601	CTGATCGAGCCCAACCCCGTGAACATCATCGGCGCGCAACATCTGACCCAGCTGGGCTGC	660	
Db	149	TTAGTAGAGACCTACACCTGTCAACATTAATTGAGAAATCTGTGACTCAGATTGG-TGC	91	
QY	661	AACCTGAACCTTCCCATCATCAGCCCATCGACGACCGTGCCCGTGAAGCTGAAGCCCGGCAATG	720	
Db	90	ACCTTAATTTTCCCATTAGCTCTATTGTAAACTGTACAGTAAATTTAAAGCAGAGAAATG	31	
QY	721	GACGCGCCCAAGGTGAAG	738	
Db	30	GATGGCCCAAAAGTAAGG	13	

RESULT 2	2886 bp	DNA	linear	GSS 21-SRP-200
CL967755				
LOCUS				
DEFINITION				
	CL967755	2886 bp	DNA	linear
	OSJFCC01518	Oryza sativa	Express Library	Oryza sativa (indica
	cultivar-group)	genomic	genomic survey	sequence.

ACCESSION	CL967755
VERSION	CL967755.1
KEYWORDS	GS8.
SOURCE	<i>Oryza sativa</i> (indica cultivar-group)
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)

REFERENCE
1 (bases 1 to 2886)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL	Unpublished (2004)
COMMENT	Contact: Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559
Fax: 86-10-80486576
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped

FEATURES	SOURCE	Location/Qualifiers
		1. 2886
		/organism="Oryza sativa (indica cultivar-group)"
		/mol_type="genomic DNA"
		/db_xref="taxon:35945"
		/clone_lib="Oryza sativa Express Library"
		/note="Oryza sativa exon tripped genomic sequences"
ORIGIN		

Query Match 4.0%; Score 98.4; DB 10; Length 2886;
Best Local Similarity 43.2%; Pred. No. 2.1e-07;
Matches 709; Conservative 0; Mismatches 906; Indels 27; Gaps 4

Db	213	GGTATATGAGAGCGCGCCGCAATATGCGCATCTGTCTCTGCGCAACGAGCGCGCGAGCGCGC	278
OY	249	GGGCAAGAGCCCGCGAGTTTCCCGACGAGCAAGAACTCGCGCAACAGCCCAACGACCCGCGA	308
Db	279	CGACTGCAAGGACTTTCGTCCGCAACGTCTGTCTCTGTATCAATCTCAACATCAATATT	338
OY	309	GTTCGACAGTTCGC---GGCGCAACCCCGCAGGAGAGCGCGAGCGCGAGCGCGAGCGAC	365
Db	339	CATGAGAGAGAACATGCTCCGGGAAACGCGCGCGCGCTCATATGCGCGAGCTCTCGCGCCAA	398
OY	366	CCTGAACCTTCCCGAGATCACTCTGTGGCAGCGCCCTGTGGTATGCAATCAAGTGGAGCG	425
Db	399	GACCAAGGTGTCTAGGAGACGGGAAATGGCAGAGACAGAGCGCTTCATCTCTGTCTCCCG	458
OY	426	CCAGATCAAGAGAGGCGCTGTGGAACACCGAGCGCGAACAACCGTGTGAGAGAGATAG	485
Db	459	CGACATCATCGACATCAAGGCTCGGCGCATCATCTCCGCGAGCGCGCGGCTGTCAAGGG	518
OY	486	CTTCGCCGCGCAATGGAAGCCCAAGATGATGAGCGCATCGAGCGGCTTCAATCAAGTTCG	545
Db	519	CGACCCGCTCAAGGTTCGACAGGCGGCGCTCAACGCGAGATCGATTCGCTCAACAGCA	578
OY	546	CCAGTATGACCAAGATCTGATTCGATCTGCGGCAAGAGGCAATCGGCAACCGTGTGAT	605
Db	579	CGCGGCGAGGAGCGCTCTTCTCCGGGTCAACGTTCAAGCAAGGACAGATTCGAGCGCTGT	638
OY	606	CGGCGCCACCCCGGTGAACATCATTCGCGCGCAATCTGACCCAGCTGGGCTTGACCT	665
Db	639	CATCGCCACCGAGCGTGACACTTCTTCGGAAGCGGCGGACCTGTGTGACAGCAACAA	698
OY	666	GAACTTCCCATCAAGCCCATCGAACCGTGGCCCGTGAAGCTGAAGCCCGGATGACGG	725
Db	699	CAACATCGGCACTTCCAGCTCGTCTCAAGGCGCATCGGCACTTCTGCAATCATCTTCAT	758
OY	726	CCCCAAGTGAAGAGTGGGCCCCCTGAACGAGAGAAATCAAGGCCCTGACCGGCACTTG	785
Db	759	CGGCGTGGCAATGATTAATCGAGATCATCTGTTATTATTCGCCATTCAGCAACCGCGGTACG	818
OY	786	CGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAACCCCTTCAACAC	845
Db	819	CGAGGGCATGACAACTGTCTGTCTGTCTCATGTGGGGGCAATCCCATTCGATGCGCCAC	878
OY	846	CCCCGTGTTGCGCATCAAGAGAGAGACAAGCAACCAATGGCGCAAGCTGTGACTTTCG	905
Db	879	CGTCTGTCCGTCACTATGCGCATCGGCTCCACCGGCTGTGCGAGAGAGGCGCATCAC	938
OY	906	CGAATCTAAACAAGGCAACCAAGACTTGTGGAGAGTGAAGTGGGCATCCCCACCCCCG	965
Db	939	CAACCGGATGACCGCATTCAGAGAGATGGCGCGCATGAGCTGTCTGCAAGCGACAAAGAC	998
OY	966	CGGCTTGAAGAGAGAGAGCGTGAACCGTGTCTGAAGTGGGCGAACCGCTTACTTCAAGCT	1025
Db	999	CGGAGCGCTGACGTTCAACAGCTCACCGTCTGAACAACGCTGATTCGAGGTGTACGGGCG	1058
OY	1026	GCCTCTGAGCAGGACTTTCGCAAGTACACCGCTTCAACCATCCCAAGCATCAACAAG	1085
Db	1059	GGGGCTGGAACAAGGACTTCGGTGTCTCTGTACGCGGAGGGGCGTCCGCGTCAAGAACCA	1118
OY	1086	GACCCCGGCACTCGGCTACCAAGTACAAACGTGTGCGCCCAAGGCTGGAAGGGCAAGCCAG	1145
Db	1119	GGAAGCAATTGACACATGATCTGTGGGCAATCTTCGCGAACCCCAAGAGAGCGCGCGCG	1178
OY	1146	CATCTTCAGAGCAGCATGACCAAGATCTGTGAAGCCCTTCCGAGCGCGCAACCCGAGAT	1205
Db	1179	CATCAAGAGGTTCACTTCTCTCCGTTCT---AACCCGTCAAGAGCGCACGAGCATAC	1235
OY	1206	CGTATCTTACAGGCGCCCGCTGTACGTGGGCAAGCACTTGAGATTCGCGACGACCGCGC	1265
Db	1236	CTACATGACCGGCAATGGCGAATGGCAACAAGATACAGCAAGGCGCGCGCGAGCAGATCAT	1295
OY	1266	CAAGATCGAGAGCTGTGCGAAGCACTGTCTGCGCTGGGGCTTCAACAACCCCGCAAGAA	1325
Db	1296	CGACTGTGTCAAGATGAGCAAGAGCGCGCGAAGAAAGATGTGACACGCTGTATCGACAGTA	1355

Oy	1326	GCACCAAGAGGCGCCCTCTTCGTGTGGAATGAGGCTGAACAGCTGCACCCCGCAAGTGCAC	1385
Db	1356	CGCGGACCGCGCGCTTCGCGTGTGCTGGCGGTGTCTGAACGAAGGTCCGAGGAAG-----	1410
Oy	1386	CGTCAGGCCCATCGAGCTGCGCCCGAAGAGAGAGCTGGAACCGTGTGAACGACATCCAGAAAGCT	1445
Db	1411	-----AGCAAGAGAGACGAGGGCGAGACCATAGGACATTTGTCGCGCTCTCTGCGGCT	1460
Oy	1446	GGTGGGCAAGCTGTAATTGAGCCAGCGCAGATCTTACCCCGGCATCAAGGTGCGCCAGCTGTG	1505
Db	1461	GTTGCACCCGCGAGGCAACGACGCGCGAGACCATTCGCGCGCGCTGCACCTTCGCGGT	1520
Oy	1506	CAAGCTGTGCGCGCGCGCCCAAGGCCCTTACCGACATCGTGGCCCTGACCCGAGAGAGCCGA	1565
Db	1521	GAACGTGAAGATGATCACCAGCGACCAAGCTGGCCATCG-----GGAAGGAGACGGCGCG	1574
Oy	1566	GCTGAGCTGACCGAGAACCGCGAGATCTGTCGCGAGCGCGGTGACCGCGCGATCTACGA	1625
Db	1575	GCGGCTCGGAGTGGCACCAACATTTACCCGTGCAGACGCTGTGCGCGACGACAGAGACG	1634
Oy	1626	CCCCGCAAGAGCCCTGTGTCGCGAGATCCAGAAGCAGGCGCACGACCACTGAGACTTACCA	1685
Db	1635	CGAGTGAAGCGGCTCTCCCATTCGACGAGCTGATCGAAGAAAGCCGACGCGCTTCGCGCGGT	1694
Oy	1686	GATCTTACCAAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGC	1745
Db	1695	GTTCGCCGAGGCAACAGTACGAGATCTGTAAAGCGCTCCAGGACCGGAAAGCACATCTGCG	1754
Oy	1746	CCACACCAACGACGTGAAGACAGCTGACCGAGGCCGTGCAGAAAGATCGCCATGAGAGCAT	1805
Db	1755	CATGACCCGCGCAGCGCGTGAACGACGCGCGCGCTGAAAGAAAGCCGACATTCGCGATCGC	1814
Oy	1806	CGTGAATCTGGGGCAAGACCCCC	1827
Db	1815	CGTCAGACGCGCACGACGCGCT	1836

RESULT 3	LOCUS	DEFINITION
CK099623	1781 bp	mRNA
CK099623		linear
		EST 03-JUN-2005
		RECC0260 A normalized whole-life-cycle cDNA library of rice <i>Oryza</i>
		sativa (indica cultivar-group) cDNA clone EI057014, THREEB09,
		EI057004, EI062D05, BI027714, EI070G16, EI08 5', mRNA sequence.

ACCESSION	CX099623
VERSION	CX099623.1
	GI:66912775

KEYWORDS	EST.
SOURCE	Oryza sativa (indica cultivar-group)

ORGANISM *Oryza sativa* (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE
AUTHORS
1 (bases 1 to 1781)
Zhang, J., Feng, Q., Jin, C., Qiu, D., Zhang, L., Xie, K., Yuan, D.,

TITLE Features of the expressed sequences revealed by a large-scale

analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63

JOURNAL
Plant J. 42 (5), 772-780 (2005)
PUBMED
15918889

COMMENT
Contact: Wang S
National Key Laboratory of Crop Genetic Improvement

Hauzhong Agricultural University
Wuhan 430070, China

Tel: 86-27-87282044
Fax: 86-27-87287092

Email: shippingwang@hotmail.com
Seq nr: 77.

FEATURES	Location/Qualifiers
Source	1-1781

/organism="Oryza sativa (indica cultivar-group)"
/mol_title="Rice"

```

/indol_type="indica"
/strain="indica"

```

```

/cultivar="Manjhi_63"
/db_xref="taxon:39946"
/clone="E1057014, THREEB09, E1057004, E1062D05, E1027114,
E1070G16, E108"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of
rice"
/notes="vector: pSPORI; Site_1: SalI; Site_2: NotI; The
library is constructed based on the strategy of saturation
hybridization with genomic DNA using rice cultivar Manjhi
63. This library consists of cDNA from 15 directionally
cloned cDNA libraries constructed with different tissues
from 9 developmental stages."

```

Query Match	3.7%;	Score 90.2;	DB 8;	Length 1781;
Best Local Similarity	43.8%;	Pred. No. 6e-06;		
Matches 447;	Conservative	0;	Mismatches 358;	Indels 6;
				Gaps 1;

231 GGACTTGCCCTCCCCAGGCAAGCCCGGAGTTCCCCAGCGAGCAGAACTGGCCAA 290

Db 194 GGACGCGGTGCTCGACGCGTGCCTCGCCAGACCCCGACAGCAGTGGCTTTCGAGAC 253

291 CAGCCCCACCGAGCCGCGAGCTTCAGAGTGCACCGCGGCGACACACCCCGCAGCGGAGGCGCGCGC 350

Db 254 GTGCACCAAGACCAATGTTGATGTTTCGGCGAGATCACCACCAAGGCCACCGTCGA 313

351 CGAGCGCCAGGGCA CCGCTGAACTTCCCCAGATCA CCGCTGTGGCA GCGCCCCCGTGGTGG 410

314 CTACGAGAGATCGTCCGCGACACTGCGCGCGCATCGGCTTCGTCGACGACGTCGG 373

411 CATCAAGCTGGGCGGCGGAGATTCAGAGAGGCGCTCTGCTGGACACCGGAGCGGAGACGACCGGT 470

[illegible][illegible][illegible]

434 GCAAGGGGAGCACGGCACATTCACCAAGGCCCCCGAGAGAGATCGGCCCCCGACCAAGG 753

511 CTTCATCAAGTGGCCAGTACGACCAGATCTGATCGAGATCTGCGCAAGGAGCCAT 530

Db 494 CCACATGTTCCGCTACGCCACGACGAGTCCCCCGAGTGTGATGCCCTCAGCCAGTCTCT 553

591 CGGACCGTCTGATCGGCCCCACCCCCGTGAATCATCGGCCGACATGCTAACCA 650

Db 554 CGCCACCAAGCTCGGCGCCCCGCTCACCGAGSTCCGCAAGAACGGCACCTGCGCGTGGCT 613

651 GCTGGGCTGCACCCCTGAACCTTCCCCCATCAGCCCCCATCGAGACCCGTGCCCCGTGAACGTGAA 710

Db 614 CAGGCCGACGGCAAGACCCAGGTCACCGTCGAGTACTTCAACGACGCCGGCGCCATGGT 673

711 GCCCGCATGACGCGCCCAAGTGAAGAGTGGCCCTGACCGAGAGAAAGATCAAGGC 770

Db 674 CCCCCTCCGCTCCACACCGTCTCATCTTCAACCGACGACGAGACCGTCACCAACGA 733

771 CCTGACCGCCATCTCCGAGAGATGAGAGGAGGCAAGATCACCAGATCGGCCCGA 830

Db 734 CGAGATCGCCGCC-----GACCTCAAGGACGTCATCAAGCCGTCATCCCGACAA 787

831 GAACCCCTACACACACCCCGTGTTCGCATCAAGAGAGGACAGCACCAAGTGGCGCAA 890

Db 788 GTACCTCGACGAGAACCATCTTCACCTCAACCCCTCGGGCCGCTTGTATCGGCGG 847

891 GCTGTGACTTCCGCGAGCTTAAACAAGCGACCCAGGACTTCTGGGAGGTGCACTGGG 950

Db 848 GCCCAGGGGACGCGCGGCTTACCGGCGGAGATCATCATGACACTTACGGGGATG 907

951 CATCCCCACCCCGCCGCTGAAGAAGAGAAGAGCGTGACCGTGCTGGACGTGGCGA 1010

Db
908 GGGCGCGACGGCGGCGGCGCTTCTCCGGCAAGAACCCGACCAAGTTCGACCGGAGCGG 967

Oy	1011	CAGCCTACTTGAAGCGCCCTTGAAAGAATTCGCGAAGTAACCGGCTTCAACATCCC	1070
Db	968	GCGCTTCACTCCGACAGGACGGCCGCNAAGCATGTGCCAGCGGCTTCGGCCGCTTG	1027
Oy	1071	CAGCATCAACAACAGACCCTCCGGCATCCGCTAACAGTCAACGTGCTGCCCGGCGCTG	1130
Db	1028	CATCGTAGAGGTGTGTGTAGGCCATCGGCGATCCCCGAGCCGCTCTCCGTGTGTCACTTC	1087
Oy	1131	GAAAGGGACGCCCCAGCATCTTTCAGAGCAGATGACAAAGATTCCTGGAGCCCTTCCGCGC	1190
Db	1088	CTACGGCACCGGCAGATCTCCCGAAGAGAGATCTCTCAAGATTCGTCAAGAGAAATTGCA	1147
Oy	1191	CCGCAACCCCGAGATCGTATCTACAGAGCCCTCCCTGTACGTGGGCAACGACTGGAGAT	1250
Db	1148	TTTTCAGGCGCCGGAGTATGACCATCAACTGAACTCAAGAGGGCGGCAACGGTTCAT	1207
Oy	1251	C 1251	
Db	1208	C 1208	

RESULT 4

LOCUS	AY103547	2598 bp	mRNA	linear	HTC 18-FEB-2005
DEFINITION	Zea mays PC0142084 mRNA sequence.				

VERSION AY103647.1 GI:21206725
KEYWORDS HTC.

ORGANISM

REFERENCES
AUTHORS
1 (bases 1 to 2598)
Gardiner, J., Schroeder, S., Polacco, M. L., Sanchez-Villeda, H.,

TITLE

JOURNAL Plant Physiol. 134 (4), 1317-1326 (2004)

REFERENCE

AUTHORS

TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of

.....

REFERENCE 3 (bases 1 to 2598)

AUTHORS

JOURNAL

COMMENT

These and other publicly available cDNAs may be found by direct searching at MSTL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

nos

```
1. "organism": "Zea mays"  
2. "mol_type": "mRNA"  
/db_xref="MAZEDB:638378"  
/db_xref="Enson:4577"  
/clone_lib="Maize Mapping Project/DuPont Cornusens  
Library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed DuPont contigs; this resource was  
assembled by DuPont as part of a collaboration for the
```

overgo addressing of BACs in conjunction with the Maize Mapping Project"

Query Match	3.5%	Score 86.8	DB 4	Length 2598
Best Local Similarity	42.1%	Pred. No. 2.5e-05		
Matches 813, Conservative	0	Mismatches 1102	Indels 15	Gaps 5

OY		296	CCACGAGCCGGAGCTGAGAGTGGCGGCACAACCCTCCGAGCGAGGCCGCGCCGAGC	355
Dd		51	CCACCACTCCCACTCTGTGCACCGCGGCCCCAACCAACACACCGCACGCGACGA	110
OY		356	GCCAGGACCCCTGAACTTCCCAGATCACCTTGTTGGCAGCGCCCTCTGTGAGCATCA	415
Dd		111	TGGCGGGCAAAGCGGCGCATCGTGGAGAAGCGAACCCTGTAACTGGGGCCGCGCGCGG	170
OY		416	AGGTGGCGCGCAATCAAGAGGCCCTGTGTGACACCGGCGCGCAACACACCGTGTGG	475
Dd		171	AGCTGGCCGGGAGCGACTGTGACGAGTGAAGCGCATGTGTGGCGTAGGCCCGGACGCCG	230
OY		476	AGGAGATGAGCTGCCCGGCAATGTGAAGGCCCAAGATATATGGCGCGCATTCGGCGCTTCA	535
Dd		231	TGTGTCAAGATGAGGGTTCCACCTTCGCGGTGGCGCAGGTGGCGCGCGTCCGCTCCGCGCA	290
OY		536	TCAAAGTGGCGCAGTACGACCAAGATCTGTATCGAATTTGGCGCAAGAGGCCATTCGCA	595
Dd		291	AGGACGCGCTCCGGCGTGGCGGTGAGCTCGACGAGAGGAGCCGCCCCCGCTCAAGGCCA	350
OY		596	CCGTGTGATGGGCCCCACCCCCGTGAACATCATTCGCGCGCAATGTCTTAACCCAGCTGG	655
Dd		351	GCACGAGTGGATCTTCACTGACATCGCCACGCGGGGGAATCTACGCGTACCAACCG	410
OY		656	GCTGCACTCTGAATCTTCCCATCAGCCCATTCGAGAAGCGTGCCTGTGAAGCTGAAGCCGG	715
Dd		411	GCTTTCGGGGGACCTTCCACCGCGGCACGAAGACGGGCCGCGCTCCAGGTCAAGCTGCG	470
OY		716	GCATGACGAGCCCAAGGTGAAGCATGAGCCCTTGACCCGAGAGGAAGATCAAGGCCCTGA	775
Dd		471	TCAAGSCATCTCAAGCGCGGAATCTTCCGCAACCGGACAGGACGAGCAACGCTGCGGTGG	530
OY		776	CCGCAATCTGGAAGAGATGGAAGAGAGGCCAAGATCAACCAATTCGGCCCCGAGAACCC	835
Dd		531	AGGTACCCCGGCGGCGAGTGTGTGTGGCATCAACCTCTTCCAGGGCTATCTCCGGA	590
OY		836	CCTACAACACCCCGGTGTTCCGCATCAAGAGAAGAGACGACGAAGTGGCGCAAGCTGG	895
Dd		591	TTCGCTTGAAGATCTCGAGGCCATCAAGAGCTGTCAACACCGGTTCAGCCCTGTCC	650
OY		896	TGAACTTTCGCGAGCTGAACAAGCGCACCCAGAGCTTCTGGAGGTGCACTGGGCAATCC	955
Dd		651	TGCGGCTTCGGGGGACCATCAACCGCGTGGCGGACCTGATCCCGCTCTCTATACGCGG	710
OY		956	CCCAACCCCGCGCGCTGAAGAAGAAAGAAAGACGATGACCGTGTGACGTGGGCGACGCT	1015
Dd		711	GCTCATATCAGGCGCGCCCAACGCGAGCGCTACCG---TCACCGGAGGAAGGTGG	767
OY		1016	ACTTCAAGGTCGCCCTGGACGAGGACTTCCGCAAGTACACCGCTTTCACCAATCCCAACA	1075
Dd		768	ACGCGCGGAGGCTTCAAGATTCGCGGCATCGAGGGCGGCTTTCAAGCTCAACCCCA	827
OY		1076	TCACAACAGAACCCCGCGGACATCCGCTACCAAGTACACGTGTGCCCCAGGCGTGGAAAG	1135
Dd		828	AGGAGGGCTTCGCATCGTCAACGGGCACGTCCGTGGGCTCCGCGCTCCGCGCACCTGTA	887
OY		1136	GCAAGCCCAAGATTTCCAGAGCAGCANTAACAAGATCTTGAAGCCCTTCCGCGCGCGCA	1195
Dd		888	TGTACAGCGCAAGTCTCTGGCGCTCTGTGAGAGTCT--GTCCGCGCTTCTTCTCG	944
OY		1196	AACCCGAGATCTGATCTTACCAAGGCCCTCCGTATCGTGGCAGCGACCTTGAGATCGGCC	1255
Dd		945	AGGTATGAAAGGCAAGCCCGAATACACGACCACTGACCCACAAGCTGAAGACACACCC	1000
OY		1256	AGCACCGCGCAAGATCGAGGAGCTCGCGCAAGCATGTGTGGCGTGGGGCTTCAACCAACC	1315

QY 541 GTGCGCCAGTACGACCAAGATCTGTATCGA---GATCTGGGCAAGAGCCATCGGACC 597
DB 691 GAGGTGCGCTACCGCGTGGGCAAGATCTCGAGATCGGCGCCAACTTCTCAACCCAGC 750
QY 598 GTGCTGATCGGCCCCCAACCCCGTGAACATCATCGGCGGCAACATCTGAACCCAGCTGGGC 657
DB 751 GTGGGATTCGGGTGCTCCGCTTCCAGAAAGCATCTCAACTGCTG----- 798
QY 658 TGGACCCCTGAATCTTCCCATTCAGCCCCCATTCGAGACCGTCCGTTGAAGCTGAAGCCCGGC 717
DB 799 TACATCTCGAGTGAATGAGCTCTCCCGAGGTGGCCAACTACTGAGAGAGGATTATCAAG 858
QY 718 ATGACGCGCCCCAAGTGAAGCATGTGCCCCCTGACCGGAGAGATCAAGCCCTGAC 777
DB 859 ATCAACGATACCAAGAGAGCGGTTCTGTAACCGGTGTGTCTGTCAATGTTCAACAC 918
QY 778 GCCATCTGGAGAGATGAGAGAGAGGCAAGATCAACAGATCGGCCCGGAGAACCC 837
DB 919 GTGGCGGCAAGAAATCGCCGTGCTGGGCTTCGCTTCAAGAGAGACACCGGCGACAC 978
QY 838 TACAACACCCCGTGTGTGCGCATCAAGAGAGAGACAGACAGATGGCGGAGCTGTGT 897
DB 979 AGGAGAGCGCCCGCATCACTGTGCAAGGCGCTGATCGGCAAGGCTCAAGGTGAGC 1038
QY 898 GACTTCGCGAGCTGAACAAGCGCACCCAGACTTCTGGAGGTGCACTGGGCACTCCC 957
DB 1039 ATCTACGACCCCGCATCACTGTGCAAGGCGCTGATCGGCAAGGCTCAAGGTGAGC 1098
QY 958 CACCCCGCGGCTTGAAGAGAGAGAGCGTGAACGCTGTCGAGCGTGGGCGACGCTAC 1017
DB 1099 GATTGGGATCAACCGGCTCATCTCCAGCGGATGAGCCCAAGGCTGATGCTC 1158
QY 1018 TTCAAGCGTGGCCCTGAGAGAGAGACTTCCGCAAGTACACGCTTCAACCTCCAGCATC 1077
DB 1159 GTGCGCTGGAGAGCTTACGAGGCGGCGGAGAGCGCCACGCGGTCTGATCTTCACTCGAG 1218
QY 1078 AACACAGAGACCCCGGCGATCGCTAC 1104
DB 1219 TGGGAGGATTCAGAGAGCTCGACTAC 1245

RESULT 6
LOCUS C2247380 743 bp DNA linear GSS 11-FEB-2005
DEFINITION A1A-asf3a12 b1 Ancylostoma caninum whole genome shotgun library
(A1AAGS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION C2247380
VERSION C2247380.1 GI:59632821
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Bukerjota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Strongylida;
Ancylostomatoidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
1 (bases 1 to 743)
Miteev,M., McCarter,J.P., Pape,D., Ritter,B., Tsagareishvili,R.,
Runko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Mesalina,D.,
Waterston,R.H., Clifton,S.W. and Wilson,R.
Genome Survey sequences from the parasitic nematode Ancylostoma
caninum
Unpublished (2004)
Contact: Miteva M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@wustl.wustl.edu
Genomic DNA provided by John Hawdon (mtmj@hgwumc.edu) DNA
sequenced by Washington University Genome Sequencing Center
Classes: shotgun.
Location/Qualifiers
1..743
/organism="Ancylostoma caninum"

/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="G510"
/clone_id="Ancylostoma caninum whole genome shotgun
library (A1AAGS 001)"
/note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
Ancylostoma caninum genomic DNA was randomly sheared,
end-repaired and size fractionated to enrich for 2-4 kb
fragments. Genomic DNA was provided by John Hawdon
(mtmj@hgwumc.edu) at George Washington University.
Sequencing by Washington University Genome Sequencing
Center, St. Louis, MO."

ORIGIN

Query Match 3.5%; Score 86; DB 10; Length 743;
Best Local Similarity 45.9%; Pred. No. 3.2e-05;
Matches 293; Conservative 0; Mismatches 345; Indels 0; Gaps 0;
QY 805 GGCAAGATCAACAGATGGGCGCCGAGAACCCCTTACACACCCCGTGTGGCATCAAG 864
DB 6 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 65
QY 865 AAGAAGACAGACCAAGTGGCGCAAGCTGTGACTTCCGCAAGTGAACAAGCGCAC 924
DB 66 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 125
QY 925 CAGGACTTCTGGAGGTGCAAGCTGGCATCCCCACCCCGCGCTGAAGAGAGAG 984
DB 126 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 185
QY 985 AGCGTACCGGTGTGAGAGTGGGCGACGCTTACGCTGACCGTCCCTGAGAGACTTC 1044
DB 186 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 245
QY 1045 CGCAAGTACACCGCTTACACCATCCCGAGCATCAACAACAACAACCCCGCATCGCTAC 1104
DB 246 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 305
QY 1105 CAGTACAAAGTGTGCGCCCAAGGCTGAAGGCGACCCCAAGATCTTCAAGAGCAGATG 1164
DB 306 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 365
QY 1165 ACCAAGATCTGTGAGCGCTTCCGCGCGGCAACCCCGAGATGTGATCTTCAAGCGCC 1224
DB 366 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 425
QY 1225 CTGTAGTGGGCGAGGACCTGAGATCGGCGGACCGCGCAAGATCGAGAGCTGCGC 1284
DB 426 AACGACAGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 485
QY 1285 AAGCACTGTGCGTGGGCTTCAACACCCCGCAAGAGAGACCAAGAGAGCGCCCC 1344
DB 486 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 545
QY 1345 TTCTGTGAGTGGGTGAAGAGTGAACCCCGCAAGTGAACCTGTGAGCGCATCGAGCTG 1404
DB 546 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 605
QY 1405 CCGGAGAGAGAGCTGAGCGGTGAACGACATCCAGAA 1442
DB 606 AACGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 643
RESULT 7
LOCUS BM321451 951 bp mRNA linear BST 03-JAN-2002
DEFINITION BM321451
Mastigamoeba balaamuthi lambda ZAP II library
3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857

Db 453 CGACCTGCG-----CGTGTTCAGCCCGAGAGAGGTGAGCGCCATGCTGCAC 500
 QY 936 GGAAGGTGACGTGGGATCCCCACCCCGCCCTGAAAGAAAGAGGTGACCGT 995
 Db 501 GGGATGAAAGAGACGGCCGAGGCTTACTCGGCAAGAGGTCAACCCGCGCTGTGAC 560
 QY 996 GCTGACGTGGGCGACGCTTACTTTCAGGCGCCCTGAGCAGAGATTCCGAAATGAC 1055
 Db 561 GGTCCCGGCTACTTTCAGAGCGCGACGGGCGACAGAGAGCGCGGCTCATCGC 620
 QY 1056 GCGCTTACCATCCCGAGATCAACAAGAGACCCCGGAGATCCGCTACAGTACAGT 1115
 Db 621 GGGGCTACCGGTGACCGCATCATCAAGAGCCCGCGCGCATGCTTACGGCAT 680
 QY 1116 GCTGCCAGGGGCTGAGAGGGGAGCCCGACGATCTTTCAGAGACATGACCAAGATCT 1175
 Db 661 CGACAAAGAGGGGCGCGAGAGAGAGTCTGTCTTTCAGCTTGGCGGCGACATTTGGA 740
 QY 1176 GAGAGCCCTTCCCGGCGCGGACCCCGAGATCGTGAATTCAGAGAGCCCGCTGTAGTG-- 1233
 Db 741 GGTGAGCATCTCGCATTCGACAGAGGCGGTTCGAGAGCTTTCGCAACCAAGCGGACAC 800
 QY 1234 ----GGCAGCGACCTGAGATGCGCCAGACCGCGCCCAAGATGAGAGCTGCGAGCA 1289
 Db 801 CCACCTCGGCGCGGAGAGCTTTCAGACCAAGCCTCATGACCACTTGTTCAGATCGC 860
 QY 1290 CCGTGTGGGCTGGGGCTTTCACACCCCGGACAAAGAGCACCAAGAGAGCCCGCTTCT 1349
 Db 861 CGGAAGCACGGCGCGGACATCAACGCGACCGCGCGCGCTGAGAGAGCTCGCGCGCA 920
 QY 1350 GTGATGAGGCTACGAGCTGACCCCGACCAAGTGAACCGTGCAGCCCATGAGCT---GCC 1406
 Db 921 GTGCGAGGGCGGCAAGCGCGGCTGACAGCAACAGACCGAGTGCGGTGAGGTGAGTGC 980
 QY 1407 CGAAGAGAGAGCTGGAACCTGACAGATCCAGAAAGCTGTGGGCAAGCTGAGCTGGC 1466
 Db 981 CCGTGTTCAGCGCGCTGACCTTGTGAGAGCGGCTTCCCGGCGCGGCTTTCAGAGAGCTCA 1040
 QY 1467 CAGCGAGATCTTACCCCGGATCAAGGTGGCCAGAGTGGCAAGCTGTGGCGGGCGGCA 1526
 Db 1041 CAGCGACCTTTCAGAAAGACATGATGTGCGGTGAGAAAGGCGCATGCGGAGGCGCGCT 1100
 QY 1527 GGCCTTGAACGACATCTGTGCGCCCTGACCGAGAGGCGGAGCTGAGAGCTGGCGAGAACCG 1586
 Db 1101 GAGCAAGGGCGGACATTCAGAGATGCTCTCGTCCGCGGACACCAAGATCCCAAGGT 1160
 QY 1587 CGAGATCTGTGCGGAGCGCGGTGCAAGAGCTGTACTACAGCCCGACAGAGAGCTGTGCG 1646
 Db 1161 GCGAGAGCTGCTCAAGAGCTACTTGGTGGCAAGAGCCCAACCGCGGCTCAACCCGCA 1220
 QY 1647 CGAGATTCAGAAAGCGGCGGACAGACAGTGAACCTTACAGATCTTACAGAGAGCCCTTCA 1706
 Db 1221 CGAGGCGGTGGGTGAGGCGCGCGGTGAGGCGCAGCATCATTCAGCGGCAAGTTCAGAGCA 1280
 QY 1707 GAACTTGAAGACCGGCAAGTACGCAAGATGCGCACCGGCGCACCAACAGAGTGAAGCA 1766
 Db 1281 GAAACAGAGAGAGATGATCTCTGAGCGTCCGCGGCTCACCTTCGCGCTGGAAGCGCG 1340
 QY 1767 GCTGACCGAGGCGGTGACAGAGATGCGATGAGAGAGATCGTATCTGGGAGAGAGCCCG 1826
 Db 1341 CGGCGGCGGTGATGCGCAAGCTGATCCCGGCAACAGGTGGTGGCGAGAGAGCGCA 1400
 QY 1827 CAAGTTCCCGCTGCCATTCAGAAAGAGACTTGGAGAGACTTGGTGAACGATCTGAGCA 1886
 Db 1401 GGTGTTTCAACAGTAAAGAGCAAGAGCAACAGGTGATTCAGAGGCGCA 1460
 QY 1887 GGCACCTTGATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGTAAGCTGTGTA 1946
 Db 1461 GGGGAGCATGAGAGAGAGCAACCGGCTGCTCGGAGGTTTCAGCTTCGCGGATGCGCC 1520
 QY 1947 CGAGTGAAGAGAGAGCCCATCATCGCGCGAGACCTTTCATGTGAGCGGCGCGCA 2006
 Db 1521 GCGCGCGAGAGGGGCGCGCGAGATGAGGTGACG-----TTTGAAGTGAAGCGCGA 1571

QY 2007 CCGGAGAGCAAGATCGGCAAGGCGGCTAGGTGACGACCGGCGGCGGAGAGATCGT 2066
 Db 1572 CGGATCTTTCAGCGGTGTGGCGCGGACAGAGCCACCGGCAAGTTCAGAGAGATCAAT 1631
 QY 2067 GAGCTGACCGAGACCAACCAACAGAGACGAGCTGAGGCGCATTCAGCTGGCCCTGCA 2126
 Db 1632 CTCGGCGGACGACCGCAAGATTCAGCAGAGAGAGTTCAGCGGATGTGGCGGAGCGGA 1691
 QY 2127 GAGACGCGGCGAGGAGTGAACATCTGTACCGACAGCATTCGCTTGGGATATCCA 2186
 Db 1692 GAGATTCCCGAGAGAGACCGCGGACAGAGAGAGGTGACGCTCCGAAACAGCTGGA 1751
 QY 2187 GGGCCGCGCGACAGAGAGAGAGGAGAGCTGTGAACAGAT-----CATGACAGCT 2240
 Db 1752 GCGTACGTTTACATCAAGAGACACGCTCGCGCGCAAGATGCGGACGCTATGAGAG 1811
 QY 2241 GATCAAGAGAGAGAGTGTACTTGAAGCTGGAGTCCCGCCCAAGAGGATTCGCGGCA 2300
 Db 1812 CGAGAGAGAGAGCAAGTGTGAGAGAGCGGTGAGAGAGGCGTACAGTGGCTGAGCGCA 1871
 QY 2301 CGAGCATTCGACAAAGCTGTGAGAGAGGCAATCCGCAAGTCTGTTCTTGAACGCA 2360
 Db 1872 CCGGACCGCGGCAAGAGAGAGTACGAGAGAGCTGAGAGAGCTGAGAGAGCTGTGCA 1931
 QY 2361 CGATGCGGAGATTCGATCTTACAGATACAGAGACGCTGACGTGAGG 2409
 Db 1932 CCGCTCATGTGCGGCTTACAGAGGTCCGCGCGGCGGCGGCG 1980

RESULT 9
 CL978463 1725 bp DNA linear GSS 21-SEP-2004
 LOCUS CL978463
 DEFINITION OsIFC031917 Oryza sativa Expressed library Oryza sativa (indica
 cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL978463
 VERSION CL978463.1 GI:52411427
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 BUKARYOTA; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 REFERENCE 1 (bases 1 to 1725)
 AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
 Zhao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
 Wong,G.K.S., Deng,X.W. and Wang,J.
 TITLE An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 JOURNAL Unpublished (2004)
 COMMENT Contact: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 FEATURES
 source
 Class: exon-trapped.
 Rice genomic sequence.
 Location/Qualifiers
 1..1725
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed library"
 /note="Oryza sativa exon trapped genomic sequences "
 ORIGIN
 Query Match 3.4%; Score 84; DB 10; Length 1725;
 Best Local Similarity 43.7%; Pred. No. 7.8e-05;
 Matches 470; Conservative 0; Mismatches 600; Indels 6; Gaps 2;
 QY 949 GGCATCCCGCACCGCGCGCTGAAAGAGAGAGGTGACCGTGTGAGCGTGGC 1008

Db 76 GTCTCCGACCCCGACGCTCTCCACCGCATGTAAGCGAGGTGAAGCTCTCTCCGAC 135
 Oy 1009 GACGCTTACTTGAAGGTGCTCTGACGAGACTTCCGAGATGACCGCTCTCAACATC 1068
 Db 136 CTCCGACGAGGCGCATATCGCTTCCAGCTCTGGGTCAACCTCTCCGCGCAC 195
 Oy 1069 CCCAGATCAACAGAGACCCCGGATCCGCTACAGTACAGTGTGCCCCAGGAC 1128
 Db 196 TTCACTTCATACAGAGCTTCTCTCCGACCTCCGCTCTTACCGCTCCGCTAC 255
 Oy 1129 TGGAGGGAGGCGCCGATCTTCCAGACAGATGACCAAGATCTGTGAGCCCTCCG 1188
 Db 256 CCGCGGTAGCGCGCGCTGCGCTGCGCGCGCGCTGCGCGCTGCGCGCTGCGCG 315
 Oy 1189 GCGCGCAACCCGAGATCTGATCTTCAAGAGCCCTCTGATGAGGAGAGAGAGAG 1248
 Db 316 TACCTTCACTCCGCGGCTCATCCAGCGGAGCTGAGTGCAGCAACATCTTCTGAC 375
 Oy 1249 ATCGGCGAGACCGCGCAAGATCGAGAGCTGCGAGACACTGCTGCGCTGGGCTTC 1308
 Db 376 GGGCACTCGGCGAGGTCAAGATCGGCACTCGGCTGCGCGCTGCTCCGCGCTCG 435
 Oy 1309 ACCACCCCGACA--AGAAGACCAAGAGAGCCCTCTCTGATGAGTGGCTACAG 1365
 Db 436 ACCCTCGCGCGAGCTCATCGGCAAGCGGAGTTGATGAGCCCGCGAGTGAACGAG 495
 Oy 1366 CTGCAACCCGAGAGTGAACCGTGAAGAGCCCATGAGCTGCGCGAGAGAGAGTGAAC 1425
 Db 496 TGTACGAGGTGAGGCGCTGATCTTCTTGGATGATGATGCTCGAGAGTCTAC 555
 Oy 1426 GTGAAGCATTCAGAGCTGTGAGCAAGCTGAGCTGAGCCAGCAATCAACCCGCG 1485
 Db 556 AACGATGACCCCTACAGCGAGTGCAGCAACCGCGCGAGATCAAGAGGTCAACGCG 615
 Oy 1486 ATCAAGGTGCGCAAGTGTGCAAGTGTGCGGCGGCGCAAGGCTCTGAGCCAGATCG-- 1543
 Db 616 GGGCACTCCCGACGCTTCTTCACTCTCAACGAGCGCGCGCTTCTGATCGG 675
 Oy 1544 -TGCCCTGACGAGAGAGCGGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1602
 Db 676 CGGTGCTTGTGACCGCGCGCGCGCGCGCTCTGCGAGAGAGTCTCTGCTGACCCCTTC 735
 Oy 1603 CCGGTGACGAGCGCTGTACTACAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1662
 Db 736 CTCTCCCGCGCTCAAGACAGAGAGATCAACAATGCTCATGCGACTGCGCGCGCA 795
 Oy 1663 GGGCAAGACAGTGAAGCTTACAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAG 1722
 Db 796 CCGCATTTGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 855
 Oy 1723 AAGTACGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1782
 Db 856 GAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 915
 Oy 1783 CAGAGATGCGCATGAGAGAGATCTGATCTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1842
 Db 916 TTCTCTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 975
 Oy 1843 ATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902
 Db 976 GTGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
 Oy 1903 GAGTGGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1962
 Db 1036 GACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1995
 Oy 1963 CCATCATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2018
 Db 1096 GGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1151

RESULT 10

Cw922203/c
 LOCUS Cw922203 1060 bp DNA linear GSS 24-FEB-2005
 DEFINITION EDCAR29YR A. castellanii, 6-8 kb library from total genomic DNA
 Acanthamoeba castellanii genomic clone EDCAR29, genomic survey
 sequence.
 ACCESSION Cw922203
 VERSION Cw922203.1 GI:60228146
 KEYWORDS GSS.
 SOURCE Acanthamoeba castellanii
 ORGANISM Acanthamoeba castellanii
 Eukaryota; Acanthamoebidae; Acanthamoeba.
 REFERENCE 1 (bases 1 to 1060)
 Anderson, I.J. and Loftus, B.J.
 Gene discovery in the Acanthamoeba castellanii genome
 JOURNAL Unpublished (2004)
 CONTACT Contact: Iain Anderson
 The Institute for Genomic Research (TIGR), www.tigr.org
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-795-7949
 Fax: 301-838-0208
 Classes: Shotgun.
 FEATURES
 source Location/Qualifiers
 1..1060
 /organism="Acanthamoeba castellanii"
 /mol_type="genomic DNA"
 /strain="Nelf"
 /db_xref="taxon:5755"
 /clone="EDCAR29"
 /clone_1b="A. castellanii, 6-8 kb library from total
 genomic DNA"
 /note="Vector: phos2"
 ORIGIN
 Query Match 3.3%; Score 80.8; DB 10; Length 1060;
 Best Local Similarity 43.9%; Pred. No. 0.00028;
 Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;
 Oy 261 CGAGTTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320
 Db 955 CTTGCGCGCGAGAGTGCACCAACTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
 Oy 321 CCGGCAACACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 380
 Db 895 CCGGCTGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836
 Oy 381 GATCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 440
 Db 835 CAGAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776
 Oy 441 CTGTGAGACACCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 500
 Db 775 GCTGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 716
 Oy 501 GAAAGCCAAAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 560
 Db 715 CCGCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 656
 Oy 561 CTTGATGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 620
 Db 655 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 596
 Oy 621 GAAATCATTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 680
 Db 595 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536
 Oy 681 CCGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 740
 Db 535 GCGCATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 476
 Oy 741 GTGGCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 797
 Db 475 CCGCTTCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 416


```

QY 1237 AGGACCTGAGATGGCGGACGACCGCGGCAAGATCGAGAGCTGGCGCAAGACCTGCTG 1296
Db 1822 GTTATGCTTCGAGCGCGCCAAAGCTCCACACCTCTCTGCGACATCCCGAGCGCG----- 1876
QY 1297 GCGTGGGGGTTTACGACCGCCCGGACAGAGAGACACAGAGAGCGCCCTTCCGCTGGATG 1356
Db 1877 -CGCGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1935
QY 1357 GCGTACGAGCTGACCGCGGACAGATGAGACCGTGCACCGCCATCGAGCTGCGCGGAGAGAG 1416
Db 1936 CGCGCGCGCGACCTTCTCTCGGTCACCGGCTGAGAGTTCTCTCGGCTGATGATCTCAAGGGC 1995
QY 1417 AGCTGACCGGTGAACGACATCCAGAGAGCTGTGTGGGCAAGCTGAATCTGGGCGACCGCATC 1476
Db 1996 GTGCGGCTGCGCGACCGGAGAGTGTGCTGCGCGAGAGATCGGCTGTGATCCACTCAG 2055
QY 1477 TACCGCGGATGAGATGGCGGACAGCTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1536
Db 2056 TACCTGTGGCTCTCGGACACACCATGAGGCACTCCCGAGCTCGGTCAAGAGGCTCGCG 2115
QY 1537 GACATGCTGCGCGCTGACCGAGAGAGCGCGAGCTGAGCTGCGCGCGGAGAGAGCGCGAGATCTG 1596
Db 2116 AACCTGCAAGACGTGAGCGTGTGGCGACAGAGCTGAGAGCGCGCTGCGCGGCTCTG 2175
QY 1597 CGCGAGCGCGGTGACGCGC 1614
Db 2176 CGCAACCCCTCGCTCGCGC 2193

```

```

RESULT 12
CK159167/c 869 bp mRNA linear EST 05-DEC-2003
LOCUS FGAS040564 Triticum aestivum FGAS: TaITS5 Triticum aestivum cDNA,
DEFINITION
RNA sequence.
ACCESSION CK159167
VERSION CK159167.1 GI:38985053
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Triticum.
1 (bases 1 to 869)
Allard, P., Crosby, M.W., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genevieve, B., Graf, R., Gulick, P., Hrycan, L.D., Laroche, A.,
Liu, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilson, D.,
Penttila, C., Roach, J.L. and Sarhan, P.
Functional Genomics of Abiotic Stress in Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.est@cs.usask.ca

```

```

FEATURES
SOURCE
Location/Qualifiers
1..869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DHS alpha"
/clone_lib="Triticum aestivum FGAS: TaITS5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression

```

```

subtractive hybridization) cDNA library from genotype
PI178383 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction,
tester) and subtracted against genotype Norstar cold
hardened at 2 C for 1 day (24 H)(driver). Modified Smart
cDNA (Clontech)priming and non-directional cloning"

```

```

Query Match 3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Pred. No. 0.0003;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

```

```

QY 580 AAGAGGCGCATGCGGACCGGTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
Db 827 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768
QY 640 ATGTGACCGGACGCTGGGCTGACCGCTGACCTTCCCATGAGCGCGCGCGCGCGCGCGCGCGCG 699
Db 767 CACAAAGAGAGACAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 708
QY 700 GTGAAGCTGAGACCGCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 759
Db 707 ACCCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 648
QY 760 AAGATCAAGGCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 819
Db 647 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 588
QY 820 ATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 879
Db 587 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 528
QY 880 AAGTGGCGGACGCTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 939
Db 527 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 468
QY 940 GTGCGAGTGGGATCCCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
Db 467 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 408
QY 1000 GACGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1059
Db 407 GACACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 348
QY 1060 TTACCATCCCGACGATCAACACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
Db 347 GACCAACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 288
QY 1120 CCCGAGGCTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1179
Db 287 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 228
QY 1180 CCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1239
Db 227 AACCAACACACACACACACACACACACACACACACACACACACACACACACACACACACACAC 168
QY 1240 GACCTGAGATCGCGCAGC 1258
Db 167 AACGCGACACACACACACAC 149

```

```

RESULT 13
CZ216254 892 bp DNA linear GSS 10-FEB-2005
LOCUS A11A-aaf23g02.b1 Ancylostoma caninum whole genome shotgun library
DEFINITION (A11AGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION CZ216254
VERSION CZ216254.1 GI:59229909
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Bukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
Ancylostomatoidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.

```


QY 507 CAAGATGATCGGCGGATCGGCGCTTCACTAAGGTCCGCAAGTACGACCAAGTCTGAT 566
 Db 405 GGCCTACCTCGGCTCGTCTCAAGAAACCCCGTGTCAACGTCGCCGCTTACTTCAACGA 464
 QY 567 CGAAGATCTGCGCAAGAGGCCATCGGCAACGCTGATCGGCCCCCAACCCCGTGAACAT 626
 Db 465 CTGCGAAGAGGCAAGGCCACCAAGAGCCCGGCTCATCGCGGCTCAAGTGAATGCGCAT 524
 QY 627 CATCGGCGCAACATGCTGACCCAGCTGGCTGACACCTGAATTCCCATCAGCCCCAT 686
 Db 525 CATCAACGAGCCACCGCGCGCATCGGTGTAACGCTCTGACCAAGAGCCGACGACGA 584
 QY 687 CGAAGCCGTGCGCTGTAAGTGAAGCCCGCATGACGCGCCCAAGGTGAAGATGCGCC 746
 Db 585 CGGCGAAGAAAGTGTCTCATTTTGAACCTCGGCGCGGCACTTGAATGTGTGCTGCT 644
 QY 747 CCGTACCGAAGAAATCAAGGCGCTGACCGCATCTGCGAAGATGGAAGAAAGGAG 806
 Db 645 CACCATCGAAGAGGAGCATCTTCAAGGTCAAGGCGACCGCGGACACCCACTCGGCGG 704
 QY 807 CAAGATCAACCAAGATGCGGCGCGGAGAACCTTAC--AACACCCCGCTGTTCCCATCA 863
 Db 705 CGAAGATCTGACCAACCGCATGTGTAACACTTCTGTGAGAGTTCAAGCGCAAGAACAA 764
 QY 864 GAAAGAAAGAACAGCAACAGTGGCGCAAGCTGTGTACTTCCGAGACTGAACAAGCGCAC 923
 Db 765 GAAAGGACATCAAGGCGCAACCCGCGCGCTCAAGGACGCGCTCAAGGACGCGCGCC 824
 QY 924 CCAAGACTTCTGTGAGATGCAAGTGGGCAATCCCCCAACCCCGCGCTGAAGAAAGAA 983
 Db 825 CAAGCGCAAGCTGTCTGTCAAGCGGCAACACCATTTGAAGTGAATCGTGTACAGAGG 884
 QY 984 GAGCGTGAACCGTCTGAGCGTGGGCGACGCGCTTACTTCAAGGTCGCCCTGACAGAGACT 1043
 Db 885 CATCGACTTCTTACACAGACATACACAGGCGCGGCTTCAAGAGACTCAACATGACCTGTT 944
 QY 1044 CGGCAAGTACACCGGCTTCAACATCCCAAGATCAACAAAGAACCCCGGCAATCCGCTA 1103
 Db 945 CGGCAAGTGCATGAGACCGCGTGAAGAGTGCCTCCGACGCGCAAGATGGAACAAGAGCAG 1004
 QY 1104 CCAAGTCAACGTGCTGCGGCGGCGTGAAGGCGACCCCAAGATCTTCAAGAGCAGCAT 1163
 Db 1005 CGTGCACGAGCTGTCTGTGCGGCGCTTCCACCGCATCCCAAGGTGAGAGCTGCT 1064
 QY 1164 GACCAAGATCTGTGAGCCCTTCCGCGCGGCAACCCGAGATGTGATCTAACAGGCCCC 1223
 Db 1065 CCAAGACTTCTTCAAGCGGAGAGGCTGTGCA-----AGAGCATCAACCCGACAGAGC 1118
 QY 1224 CCTGTACGTGGGCGAGCACTGTGAGATCGGCGCAGCACCGGCGCAAGATCGAGAGCTGCG 1283
 Db 1119 CGTGGGCTTCAAGCGGCGGCTTCAAGGCGGCGCATCTGAGGCGGAGAGGCAAGAGGT 1178
 QY 1284 CAAGCACTGTGCTGCGTGGGCTTCAACACCCCGGCAAGAACACCAAGAGAGGCCCC 1343
 Db 1179 GCGAGCACTGTCTCTGTGACGTCAACCCGCGTGTCTTCCCTGCGCTGAGAGCCCGCGG 1238
 QY 1344 CTTCCTGTGAGATGGGCTGAGAGCTGACCCCGCAAGATGAGACCTGCA---GCCCATCA 1400
 Db 1239 CGTCAATGACCGTCTCTATCCGAGAACACACATCCGACCAAGAGAGAGGCTTT 1298
 QY 1401 GCTGCCCGAAGAGAGAGTGAACCGTGAACGATCCAGAACTGTGTGGCAAGCTGA 1460
 Db 1299 CTCACACTTACTCGACAAACAGCCCGCGCTGATCTAAGTGAAGAGGCGAGCGCG 1358
 QY 1461 CTGGGCGAGCAAGATCTAACCCCGGATCAAGGTCGCGCACTGTGCA-----AGCT 1511
 Db 1359 CAGAGACCAAGAGAACAACTCTCGGCAAGTTGAGACTTCCGCGCATCCACCGCGCC 1418
 QY 1512 GCTGCGGCGGCGCAAGGCGCTGAACCGATCGTGTGCGCTGACGAGAGGCGGAGCTGA 1571
 Db 1419 GCGCGGCGTCCCAAGTCAACGTGTCTTCAAGTGAACGCGCAACGCGCATCTTGAACGT 1478

QY 1572 GCTGGCGAAGAACCGGAGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1631
 Db 1479 GTCCGGAGAGGACAAAGACGCGGCGAAGAAACAAATACATCAACCAAGACAAAGG 1538
 QY 1632 CAAGACCTGTGTGGCGGAGATCAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1691
 Db 1539 GCGGCTGAGCAAGAGAGAGATGAGAGATGTGTGAGAGGCGGCGGAGATCAAGGCGG 1598
 QY 1692 CCAAGAGCCCTTCAAGAACCTTAAGACCGGCAAGTACGCGCAAGATGCGGCGGCGG 1751
 Db 1599 GAGCGAGAGGACCAAGAAAGATGTGACCGCAAGAGCGGCTTCAAGATCAAGGCTTCA 1658
 QY 1752 CAACGAGTGAAGAGCTGACCGAGGCGGTGAGAAAGATGCGCATGAGAGATGCTGAT 1811
 Db 1659 CATGCGCAACACATCAAGAGAGAGAGATGCGCTTCAAGCTGCGCGGCGGCAAGAA 1718
 QY 1812 CTGGGCGAAGACCCCAAGTTCCGCTGCGCTCAACGAGAGAGACTTGAAGACTGTGTG 1871
 Db 1719 GAGATTCAGAGAGCGCATGATGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1778
 QY 1872 GACCGACTTGTGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1931
 Db 1779 GCGCGAGAGTTCAGAGACA--AGATGAAGAGGCTGAGGCGGCGGCGGCGGCGGCGG 1835
 QY 1932 GGTGAAGCTGTGTGTACAGCTGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1991
 Db 1836 CCGCAAGATGTACAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1895
 QY 1992 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2036
 Db 1896 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1940

RESULT 15
 BM320864
 LOCUS
 DEFINITION
 Masticamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
 sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Masticamoeba balamuthi
 Masticamoeba balamuthi
 Bukariyoca; Pelobiontida; Masticamoebidae; Masticamoeba.
 REFERENCE
 1 (bases 1 to 1132)
 Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,
 Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
 Philippe, H.
 The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Masticamoeba
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 COMMENT
 11830664
 Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockefeller.edu
 Insert Length: 1132 Std Error: 0.00
 POLYA=No.

FEATURES
 source
 1..1132
 /organism="Masticamoeba balamuthi"
 /mol_type="mRNA"
 /strain="ATCC 30984"
 /db_xref="taxon:108607"
 /clone_id="Masticamoeba balamuthi lambda ZAP II library"
 /note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.3%; Score 80.2; DB 3; Length 1132;
 Best Local Similarity 45.3%; Pred. No. 0.00037;
 Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

QY 139 CGCGCCCCCGAAGAGGCTGTGAAGTCCGCAAGAGGCGCACCAATGAAGAC 198
DB 79 CGCCCGCAGAGGCAAGCGGCTACCGCGCGCGCCANCTGTATCAGAGCAAGAACAG 138
QY 199 TGCACCGAGCGCAGGCGCACTTCTTCCGCGAGAGACTGCGCTTCCCGCAGGCGCAAGCC 258
DB 139 TACCAACAGCCCAAGTACCGCTTCGTC-----GTCCGCTTCAACCAACAGGAGCATCGTC 192
QY 259 CGCGAGTTCCCGAGGAGAGAACCGCGCCCAAGCCCCCAAGCGCGAGCTGCAGGTG 318
DB 193 TGCAGATGCGCTTACCGCAAGATCGACGCGAGCACATCTGCGCGCGCTACTCGCAC 252
QY 319 CGCGGCGAACAACCCCGCAGCGAGCGCGCGCCAGCGCAACCTGAATTCCCGC 378
DB 253 GAGCTCACCGCTTCGCGGTCAAGCTCGGCTGACCACTACGCGCGCTTACGGACT 312
QY 379 CAGATCACCTGTGTGCAAGCGCCCTGTGTGAGATCAAGTGGCGGCGCATCAAGAG 438
DB 313 GGCCTGTGCTGCGCGCGCGTGTGTGAAGAACTCAACTCGACTCCAAGTACGAGGT 372
QY 439 GCGCTGTGTGAACACCGCGCGCGCAACCGATGTGAAGAGATGAGCTGCGCGGCAAG 498
DB 373 GTCAAGAAAGGTCAACGCGAGGACTTACACGTGAGAGAGCTGACGACGCGCGCGGT 432
QY 499 TGGAAAGCCCAAGATGATCGGCGCGCATCGCGCGCTTATCAAGGTGCGCGATACGACAG 558
DB 433 TTCAAGGCGCTGTCTGACGTGCGCGCTGTCTGACCTGACGTGCGCGCGCGGTTCGCC 492
QY 559 ATCTGATGAGATCTGCGCGCAAGAGGCGCATCGGCACTGTGTGATGCGCGCGCACCGCC 618
DB 493 GCGCTCAAGGCGATGTGCGACGCGCGGTCAAGTCCCG---CACAGCGAGACCGGCTTC 549
QY 619 GTGAACATCATGTGCGCGCAACATGTGACCCGAGTGGCTGCAACCTTCCCATC 678
DB 550 GTGCGCTTCAACGCGCGCAACAGAGAGCTCAACGCGCGGTCTCTCGCAATCATCTTC 609
QY 679 AGCCCATTCGAGACGCTGCGCGTGAAGCTGAAGCCCGCATGAGCGCGCCCAAGGTGAAG 738
DB 610 GCGCGCGCACGTGCGCGGTATCATGAAGCTCTCAAGAGACAGGACGCGCGCGCTTCGAC 669
QY 739 CAGTGGCCCTTGAACGAGAGAGATCAAGCGCTGACCGCACTTGCAGAGAGATGAG 798
DB 670 CGCCAGTTCTCGCGCTACCGCAAGAGGAGGTGTACCGCGCATGTCTGAGAAATCTAC 729
QY 799 AAGAGGCGCAATACCAAGATCGGCGCGCGCAACCCCTACACACCGCGGTTCGCC 858
DB 730 ACCGAGGCGCCACAGCAAGATCGCGCGGACCGGACTTGTCCCAAGCGCGCTCGAAG 789
QY 859 ATCAAGAGAGAGACAGCACCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAG 918
DB 790 CCGAGGCGCGCAAGCGCAAGCATGTGGGCAAGCGGCTGACGTACACAGAGCGCAAG 849
QY 919 CGCACCCAGGACTTCTGAGGAGGTGACGTGGGATCTCCCGACCGCGCC 966
DB 850 AACCGGTGCGCCAGAGAAAGTCCGCTGGGCTACCCGACGCGCCC 897

Search completed: December 31, 2005, 02:31:26
Job time : 8703.67 secs

Db	181	GGCCACCAAGATGAGAGACTGGACCGAGGCGCAAGCCACTTCTTCGAGAGACTGGCC	240
Oy	241	TTCCCCCAGGGGCAAGGCGCGAGTTTCCCAAGCGACAGAAACCGGCGCAAGGCCAC	300
Db	241	TTCCCCCAGGGGCAAGGCGCGAGTTTCCCAAGCGAGAGAAACCGGCGCAAGGCCAC	300
Oy	301	AGCCGCGAGCTGCAAGTGTGCGGCGCAACCCCGCAAGAGGCTGGGCGCGAGGCGAG	360
Db	301	AGCCGCGAGCTGCAAGTGTGCGGCGCAACCCCGCAAGAGGCTGGGCGCGAGGCGAG	360
Oy	361	GGCACCCTGAACTTCCCGCAGATCAACCTGTGTGACAGCGCCCGTGTGATCAAGGTG	420
Db	361	GGCACCCTGAACTTCCCGCAGATCAACCTGTGTGACAGCGCCCGTGTGATCAAGGTG	420
Oy	421	GGCGGCGCAGATCAAGAGAGCGCTGTGACACCGGCGCGACACACCGTGTGAGAG	480
Db	421	GGCGGCGCAGATCAAGAGAGCGCGCTGTGACACCGGCGCGACACCGTGTGAGAG	480
Oy	481	ATGAGCCTTGCCTCGGCAAGTGGAGGCCCAAGATGATCGCGGCGATCGCGCTTCAAC	540
Db	481	ATGAGCCTTGCCTCGGCAAGTGGAGGCCCAAGATGATCGCGGCGATCGCGCTTCAAC	540
Oy	541	GTGCGCGCAGTACGACACATCTGTATCGAGATCTGTGGGCGAGAAAGGCCATCGGCG	600
Db	541	GTGCGCGCAGTACGACACATCTGTATCGAGATCTGTGGGCGAGAAAGGCCATCGGCG	600
Oy	601	CTGATCGGCGCCCAACCCCGTGACATCATCTGGCGCAATGCTGACCCAGCTGGCTGC	660
Db	601	CTGATCGGCGCCCAACCCCGTGACATCATCTGGCGCAATGCTGACCCAGCTGGCTGC	660
Oy	661	ACCTGAACTTCCCATCATAGCCCATCGACACCGTGCCTGTAAAGCTAAAGCCCGGCA	720
Db	661	ACCTGAACTTCCCATCATAGCCCATCGACACCGTGCCTGTAAAGCTAAAGCCCGGCA	720
Oy	721	GACGGCGCCCAAGGTGAGACAGTGGCGCCCTGACCGAGAGAAAGATCAAGGCTTGACG	780
Db	721	GACGGCGCCCAAGGTGAGAGTGGCGCCCTGACCGAGAGAAAGATCAAGGCTTGACG	780
Oy	781	ATCTGCGAGAGATGAGAGAGAGGCGCAAGTCAACCAAGATCGGCGCGAGAACCTCTAC	840
Db	781	ATCTGCGAGAGATGAGAGAGAGGCGCAAGTCAACCAAGATCGGCGCGAGAACCTCTAC	840
Oy	841	AAACACCCCGTGTTCGCATCAAGAGAGAGCAGCACCAAGTGGCGCAAGCTGTGAG	900
Db	841	AAACACCCCGTGTTCGCATCAAGAGAGAGCAGCACCAAGTGGCGCAAGCTGTGAG	900
Oy	901	TTTCGCGAGCTGGAACAAGCGCACCCAGGACTTCTGTGAGAGTGTGCACTGGGCATCC	960
Db	901	TTTCGCGAGCTGGAACAAGCGCACCCAGGACTTCTGTGAGAGTGTGCACTGGGCATCC	960
Oy	961	CCCGCGGCGCTGGAAGAAGAAGAAGCGTGAACGTGTGAGACGTGGGCGACGCTTACT	1020
Db	961	CCCGCGGCGCTGGAAGAAGAAGAAGCGTGAACGTGTGAGACGTGGGCGACGCTTACT	1020
Oy	1021	AGCGTGCCTGTGAGCAGAGCTTCCGCAAGTACACCGCTTCAACATCCAGGATCAAC	1080
Db	1021	AGCGTGCCTGTGAGCAGAGCTTCCGCAAGTACACCGCTTCAACATCCAGGATCAAC	1080
Oy	1081	AACGAGACCCCGGCGATCGCTACCAAGTACAACTGTGCGCCCAAGGCTGGAAGGCGAG	1140
Db	1081	AACGAGACCCCGGCGATCGCTACCAAGTACAACTGTGCGCCCAAGGCTGGAAGGCGAG	1140
Oy	1141	CCGACAGATCTTCAGAGCAGACGTACAAAGATCTGTGAGCGCCCTTCGCGCGCGCAAC	1200
Db	1141	CCGACAGATCTTCAGAGCAGACGTACAAAGATCTGTGAGCGCCCTTCGCGCGCGCAAC	1200
Oy	1201	GAGATCTGATCTTACAGGCGCCCGCTGTATCTGTGGCGAGCACTTGAGATCGGCGCAG	1260
Db	1201	GAGATCTGATCTTACAGGCGCCCGCTGTATCTGTGGCGAGCACTTGAGATCGGCGCAG	1260
Oy	1261	CGCGCAAGATGAGAGAGTGTGCGACGACCTGTGCGTGTGGGCTTCAACACCCCGAC	1320

D	b	1261	CGCGCCAAAGATCGAAGAGACGTGCGCAAGACCTGTGTGGCTGTGGGGCTTCAACACCCCGCAG	1320
O	y	1321	AAGAAGCACCAAGAAAGAGCCGCCCTTCTGTGCCATCGAGCTGCACCCCGACAAGATGAGCC	1380
D	b	1321	AAGAAGCACCAAGAAAGAGAGCCGCCCTTCTGTGCCATCGAGCTGCACCCCGACAAGATGAGCC	1380
O	y	1381	GTGCAGGCCCATTCGAGCTGTGCCGAGAAAGAAAGCTGTGACCGTTGAACCGACATTCGAAAGCTG	1440
D	b	1381	GTGCAGGCCCATTCGAGCTGTGCCGAGAAAGAAAGCTGTGACCGTTGAACCGACATTCGAAAGCTG	1440
O	y	1441	GTGGGCAAGCTGAACCTGGGCCAGCCAGCATGTACCCCGGACATCAAGGTGGCCGCAAGCTGTGC	1500
D	b	1441	GTGGGCAAGCTGAACCTGGGCCAGCCAGCATGTACCCCGGACATCAAGGTGGCCGCAAGCTGTGC	1500
O	y	1501	AAGCTGTGTGCGCGCGCCCAAGGCCCTTGACCGGACATGTGTGCCCTTGACCGAGAGAGCCGAG	1560
D	b	1501	AAGCTGTGTGCGCGCGCCCAAGGCCCTTGACCGGACATGTGTGCCCTTGACCGAGAGAGCCGAG	1560
O	y	1561	CTGGAAGCTGTGCGCGAGAACCCGAGAGATCTGTGCGGAGGCCCTGTGCACCGCGCTGTACTAGAC	1620
D	b	1561	CTGGAAGCTGTGCGCGAGAACCCGAGAGATCTGTGCGGAGGCCCTGTGCACCGCGCTGTACTAGAC	1620
O	y	1621	CCGAGCAAGGACCGTGTGTGCGGAGATCCAGAAAGAGGGCCACCGACAGTGTGACCTTACAG	1680
D	b	1621	CCGAGCAAGGACCGTGTGTGCGGAGATCCAGAAAGAGGGCCACCGACAGTGTGACCTTACAG	1680
O	y	1681	ATCTTACCAAGAGGCCCTTCAAGAACTTGAAGACCGGACAGTACCGCAAGATGTGCGACCGGC	1740
D	b	1681	ATCTTACCAAGAGGCCCTTCAAGAACTTGAAGACCGGACAGTACCGCAAGATGTGCGACCGGC	1740
O	y	1741	CACACCAACCGACGTGAAGCAGCTGACCGAGGCCGTGTGAAAGATGTGCATGTGAGAGCATC	1800
D	b	1741	CACACCAACCGACGTGAAGCAGCTGACCGAGGCCGTGTGAAAGATGTGCATGTGAGAGCATC	1800
O	y	1801	GTGATCTGTGGGCCAAGACCCCCCAAGTTTCGCGCTGCGCATCCAGAAAGAGACCTGTGAGAACCC	1860
D	b	1801	GTGATCTGTGGGCCAAGACCCCCCAAGTTTCGCGCTGCGCATCCAGAAAGAGACCTGTGAGAACCC	1860
O	y	1861	TGTGTGACCCGACCTACTGTGGCAGGCGCACCTGTGATCCCGAGTGTGGAGTTCTGTGAACACCCC	1920
D	b	1861	TGTGTGACCCGACCTACTGTGGCAGGCGCACCTGTGATCCCGAGTGTGGAGTTCTGTGAACACCCC	1920
O	y	1921	CCCTCTGTGAAGCTGTGTGTACCAAGCTGTGAGAAAGAGGCCCATCATCGGCGCCGAGACTTTC	1980
D	b	1921	CCCTCTGTGAAGCTGTGTGTACCAAGCTGTGAGAAAGAGGCCCATCATCGGCGCCGAGACTTTC	1980
O	y	1981	TACCTGTGACCGGCGCGCGCCCAACCGCGGAGACCAAGATGTGGACAAAGCGCGCTTACGTGACCGAC	2040
D	b	1981	TACCTGTGACCGGCGCGCGCCCAACCGCGGAGACCAAGATGTGGACAAAGCGCGCTTACGTGACCGAC	2040
O	y	2041	CGGGGCGGGCGAGAAAGATCTGTGAGCGCTTACCGAGACCAACCAACCGAAGAACCGAGCTTGAC	2100
D	b	2041	CGGGGCGGGCGAGAAAGATCTGTGAGCGCTTACCGAGACCAACCAACCGAAGAACCGAGCTTGAC	2100
O	y	2101	GCCATCCAGCTGTGCGCTTGACAGACCAAGCGGACCGAGGTTGAACATCTGTACCGACGCGACG	2160
D	b	2101	GCCATCCAGCTGTGCGCTTGACAGACCAAGCGGACCGAGGTTGAACATCTGTACCGACGCGACG	2160
O	y	2161	TACGCCCTCTGGGACATCATTCAGGCCCAACCCCGACAAGACCGAGACCGAGCTGTGTAAACAG	2220
D	b	2161	TACGCCCTCTGGGACATCATTCAGGCCCAACCCCGACAAGACCGAGACCGAGCTGTGTAAACAG	2220
O	y	2221	ATTCATTCGAGCAGCTGTATCTAAGAAAGAAAGGTATCTCTGAGCTGTGGGTGCCCGGCCACAAG	2280
D	b	2221	ATTCATTCGAGCAGCTGTATCTAAGAAAGAAAGGTATCTCTGAGCTGTGGGTGCCCGGCCACAAG	2280
O	y	2281	GGCATTCGGCGGCAACGAGCAGATGTGACAAAGCTGTGTGAGCAAGGGCATTCGCAAGGTGTCTG	2340
D	b	2281	GGCATTCGGCGGCAACGAGCAGATGTGACAAAGCTGTGTGAGCAAGGGCATTCGCAAGGTGTCTG	2340
O	y	2341	TTCTCTGAGCGGCAATCGATGTGCGCGGACATGTGTATCTTACCAAGTACATGAGCAACCTGTACTGT	2400
D	b	2341	TTCTCTGAGCGGCAATCGATGTGCGCGGACATGTGTATCTTACCAAGTACATGAGCAACCTGTACTGT	2400

Query 2401 GGCAGCGGGGCGCCCTAGATCCGATTTAAAGCTTCCCGGGGCTTAGCACCAGTGGAATTC 2457
Db 2401 GGCAGCGGGGCGCCCTAGATCCGATTTAAAGCTTCCCGGGGCTTAGCACCAGTGGAATTC 2457

RESULT 2
AX455915
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS
SOURCE
ORGANISM
1. synthetic construct
synthetic construct
other sequences; artificial sequences.

REFERENCE
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source 1. 2463
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN
Query Match 99.2%; Score 2436.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 5.5e-201;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

1 GTGCAAGCCCACTGAGCGGCGGCGCATGAGCCAGCCAGCCAGCCCAATCTTGATGAG 60
1 GTGCAAGCCCACTGAGCGGCGGCGCATGAGCCAGCCAGCCAGCCCAATCTTGATGAG 60
61 CGCAGCACTTCAAGGAGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGCG 120
61 CGCAGCACTTCAAGGAGCGCCCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGCG 120
121 CACATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
121 CACATGCGCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
181 GGCACACAGATGAAGGACTGCAACGAGCGCGAGCGCAACTTCTTCCGCGAGGACTGGCC 240
181 GGCACACAGATGAAGGACTGCAACGAGCGCGAGCGCAACTTCTTCCGCGAGGACTGGCC 240
241 TTCCCGCAGGCGCAAGGCGCGCGAGTTCCCGAGGAGCAAGCCGCGCAACAGCCCGAC 300
241 TTCCCGCAGGCGCAAGGCGCGCGAGTTCCCGAGGAGCAAGCCGCGCAACAGCCCGAC 300
301 AGCGCGAGCTGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
301 AGCGCGAGCTGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
361 GGCACCTGAACCTTCCCGCGAGTCAACCTGTGCGAGGCGCGCGCGCGCGCGCGCGCGCG 420
361 GGCACCTGAACCTTCCCGCGAGTCAACCTGTGCGAGGCGCGCGCGCGCGCGCGCGCGCG 420
421 GCGCGCAGATCAAGAGGCGCGCTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
421 GCGCGCAGATCAAGAGGCGCGCTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
481 ATGAGCTGCTGCGCGCAAGTGAAGCCCAAGATGATGGCGCGCATCGCGGCTTCACTCAAG 540
481 ATGAGCTGCTGCGCGCAAGTGAAGCCCAAGATGATGGCGCGCATCGCGGCTTCACTCAAG 540
541 GTGCGCAAGTGAAGCAAGATCTGATGAGATCTGCGCGCAAGAGGCGCATCGCGCGCGCG 600
541 GTGCGCAAGTGAAGCAAGATCTGATGAGATCTGCGCGCAAGAGGCGCATCGCGCGCGCG 600

Query 601 CTGATGCGCCCAACCCCGCTGAACATCATGCGCGCGCAAGATGCTGACCCGAGCTGGCTGC 660
Db 601 CTGATGCGCCCAACCCCGCTGAACATCATGCGCGCGCAAGATGCTGACCCGAGCTGGCTGC 660
661 ACCCTGAACCTTCCCGCATGAGCCCGCATGAGACCGGTGCGCGGTGAAGCTGAAGCCCGCAT 720
661 ACCCTGAACCTTCCCGCATGAGCCCGCATGAGACCGGTGCGCGGTGAAGCTGAAGCCCGCAT 720
721 GACGCGCCCAAGTGAAGAGTGGCGCGCTGACCGGAGAGAGATCAAGGCGCTGACCGCC 780
721 GACGCGCCCAAGTGAAGAGTGGCGCGCTGACCGGAGAGAGATCAAGGCGCTGACCGCC 780
781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAATGCGCGCGCGCGCGCGCGCG 840
781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACCAATGCGCGCGCGCGCGCGCGCG 840
841 AACACCCCGGTGTTGCGCATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
841 AACACCCCGGTGTTGCGCATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
901 TTCCGCGAGCTGAACAGGCGCACCCAGAGCTTCTGAGAGGTGAGCTGCGCATCCCGAC 960
901 TTCCGCGAGCTGAACAGGCGCACCCAGAGCTTCTGAGAGGTGAGCTGCGCATCCCGAC 960
961 CCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
961 CCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
1021 AGCGTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1021 AGCGTGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
1081 AACGAGACCCCGGCGCATCGCTACAGAGTCAAGAGTGTGCGCGCGCGCGCGCGCGCG 1140
1081 AACGAGACCCCGGCGCATCGCTACAGAGTCAAGAGTGTGCGCGCGCGCGCGCGCGCG 1140
1141 CCGAGCATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1141 CCGAGCATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1201 GAGATGATGATCAACAGGCGCGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1201 GAGATGATGATCAACAGGCGCGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1201 GAGATGATGATCAACAGGCGCGCGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
1261 CCGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1261 CCGCGCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
1321 AAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
1321 AAGAGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
1375 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
1375 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
1381 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
1381 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
1435 AAGCTGTGGGCAAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
1435 AAGCTGTGGGCAAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
1441 AACCTGTGGGCAAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1441 AACCTGTGGGCAAGCTGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1495 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGCTGCGCTGACCGAGAG 1554
1495 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGCTGCGCTGACCGAGAG 1554
1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGCTGCGCTGACCGAGAG 1560
1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGCTGCGCTGACCGAGAG 1560
1555 GCGGAGCTGAAGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
1555 GCGGAGCTGAAGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
1561 GCGGAGCTGAAGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
1561 GCGGAGCTGAAGCTGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
1615 TACGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
1615 TACGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
1621 TACGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
1621 TACGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680

QY 1675 TACGAGATCTACGAGAGCCCTTCAAGACTGAAGACCGGACAGTACGCCAAGATGCGC 1734
 DB 1681 TACGAGATCTACGAGAGCCCTTCAAGACTGAAGACCGGACAGTACGCCAAGATGCGC 1740
 QY 1735 ACCGCGCACACCAAGAGCTGAAGAGCTGACCGAGCCCTGCAAGAGATCGCCATGAG 1794
 DB 1741 ACCGCGCACACCAAGAGCTGAAGAGCTGACCGAGCCCTGCAAGAGATCGCCATGAG 1800
 QY 1795 AGCATGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCGCATCCAGAGAGACTG 1854
 DB 1801 AGCATGTGATCTGGGGCAAGACCCCAAGTTCGGCTGCGCATCCAGAGAGACTG 1860
 QY 1855 GAGACCTGTGAGACCACTGAGAGGCACTGATCCCGAGTGGAGTTCTGTAA 1914
 DB 1861 GAGACCTGTGAGACCACTGAGAGGCACTGATCCCGAGTGGAGTTCTGTAA 1920
 QY 1915 ACCCCCCCTGTGTGAAGTTGTGTGACCAAGCTGAGAGAGAGCCCATCTGCGCCGAG 1974
 DB 1921 ACCCCCCCTGTGTGAAGTTGTGTGACCAAGCTGAGAGAGAGCCCATCTGCGCCGAG 1980
 QY 1975 ACCCTTACGTGAGAGCGGCGGCAACCGGAGACCAAGATGCGGAGAGCGGCTAAGT 2034
 DB 1981 ACCCTTACGTGAGAGCGGCGGCAACCGGAGACCAAGATGCGGAGAGCGGCTAAGT 2040
 QY 2035 ACCGACCGGCGCGGCAAGAGATGCTGAGCTGACCGAGACCAACCAAGAGACCGAG 2094
 DB 2041 ACCGACCGGCGCGGCAAGAGATGCTGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
 QY 2095 CTGCAAGGCGATCAAGCTGCGCTGTGAGAGAGCGGAGCGGAGTGAACATCTGTACCGAC 2154
 DB 2101 CTGCAAGGCGATCAAGCTGCGCTGTGAGAGAGCGGAGCGGAGTGAACATCTGTACCGAC 2160
 QY 2155 AGCGATACGCGCTGGGCGATCATCCAGGCGGCGGCAAGAGAGAGAGCGGCTGAGT 2214
 DB 2161 AGCGATACGCGCTGGGCGATCATCCAGGCGGCGGCAAGAGAGAGAGCGGCTGAGT 2220
 QY 2215 AACGAGATCATGAGAGAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGTGGCCGCG 2274
 DB 2221 AACGAGATCATGAGAGAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGTGGCCGCG 2280
 QY 2275 CACAGAGGCGATGCGGCGCAAGAGATGCAAGAGCTGTGTAGAGAGAGCGATTCGCGAG 2334
 DB 2281 CACAGAGGCGATGCGGCGCAAGAGATGCAAGAGCTGTGTAGAGAGAGCGATTCGCGAG 2340
 QY 2335 GTGCTGTTCTGGAAGGAGATGATGCGGCGCATGATATACAGATCACTGAACGACCTG 2394
 DB 2341 GTGCTGTTCTGGAAGGAGATGATGCGGCGCATGATATACAGATCACTGAACGACCTG 2400
 QY 2395 TACGTGAGGAGCGGCGGCTTAGATTCATTAAGCTTCCGCGGCTTAGACCGGAGTA 2454
 DB 2401 TACGTGAGGAGCGGCGGCTTAGATTCATTAAGCTTCCGCGGCTTAGACCGGAGTA 2460
 QY 2455 TTC 2457
 DB 2461 TTC 2463

RESULT 3
 AX455914 2469 bp DNA linear PART 06-JUL-2002
 LOCUS AX455914
 DEFINITION Sequence 30 from Patent WO0204493.
 ACCESSION AX455914
 VERSION AX455914.1 GI:21714899
 KEYWORDS
 ORGANISM
 SOURCE
 other sequences; artificial sequences.
 REFERENCE
 1 zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, B.
 AUTHORS
 TITLE
 JOURNAL
 Chiron Corporation (US) ; University of Stellenbosch (ZA)

FEATURES
 source
 Location/Qualifiers
 1..2469
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="PR975(+)"

ORIGIN
 Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
 Best Local Similarity 99.3%; Pred. No. 3.3e-199;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 1 GTGAGCGCACCATGCGGAGGCGCATGAGCGGACCAAGCGGCAAGACTCTGTAGAG 60
 DB 1 GTGAGCGCACCATGCGGAGGCGCATGAGCGGACCAAGCGGCAAGACTCTGTAGAG 60
 QY 61 CGGAGCACTTCAGAGGCGCCCAAGCGCATCATAGTCTTCACTGCGGCAAGAGGCG 120
 DB 61 CGGAGCACTTCAGAGGCGCCCAAGCGCATCATAGTCTTCACTGCGGCAAGAGGCG 120
 QY 121 CACATGCGCGGCAACTGCGCGGCGCGCGCAAGAGGCGTGTGAGAGTGGCGAAGAG 180
 DB 121 CACATGCGCGGCAACTGCGCGGCGCGCGCAAGAGGCGTGTGAGAGTGGCGAAGAG 180
 QY 181 GCGCACAGATGAGAGACTGACACGAGCGGCGCAACTTCTTCGCGAGAGACTTGGCC 240
 DB 181 GCGCACAGATGAGAGACTGACACGAGCGGCGCAACTTCTTCGCGAGAGACTTGGCC 240
 QY 241 TTCCCCAGGCGAGAGGCGCGAGTTTCCCGAGCGAGCAAGACCGGCGCAAGCCCGAC 300
 DB 241 TTCCCCAGGCGAGAGGCGCGAGTTTCCCGAGCGAGCAAGACCGGCGCAAGCCCGAC 300
 QY 301 AGCGCGAGCTGAGAGTGGCGGCGAGCAAGCCCGGAGAGAGGCGGCGGAGGCGAG 360
 DB 301 AGCGCGAGCTGAGAGTGGCGGCGAGCAAGCCCGGAGAGAGGCGGCGGAGGCGAG 360
 QY 361 GCGACCTTGAACCTTCCCGCATGATCACTGTGTGAGAGCGCCCTCTGTAGAGACTG 420
 DB 361 GCGACCTTGAACCTTCCCGCATGATCACTGTGTGAGAGCGCCCTCTGTAGAGACTG 420
 QY 421 GCGGCGAGATCAAGAGAGGCGCTGTGAGACCGGCGCGGAGACCGGTGTGAGAG 480
 DB 421 GCGGCGAGATCAAGAGAGGCGCTGTGAGACCGGCGCGGAGACCGGTGTGAGAG 480
 QY 481 ATGAGCGTGGCGGAGAGTGAAGGCGCAAGTATGCGCGGATGCGGCGCTTCAATCA 540
 DB 481 ATGAGCGTGGCGGAGAGTGAAGGCGCAAGTATGCGCGGATGCGGCGCTTCAATCA 540
 QY 541 GTGCGCGAGTACGACCAAGTCTGATTCGAGATCTTGGCGCAAGAGCGCATTCGGAC 600
 DB 541 GTGCGCGAGTACGACCAAGTCTGATTCGAGATCTTGGCGCAAGAGCGCATTCGGAC 600
 QY 601 CTGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 DB 601 CTGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
 QY 661 ACCCTGAACCTTCCCGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 DB 661 ACCCTGAACCTTCCCGCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
 QY 721 GACGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 DB 721 GACGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 QY 781 ATCTGCGAGAGATGAGAGAGAGAGAGAGAGATCAAGATGCGGCGGCGGAGAG 840
 DB 781 ATCTGCGAGAGATGAGAGAGAGAGAGAGAGATCAAGATGCGGCGGCGGAGAG 840
 QY 841 AACACCGCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 DB 841 AACACCGCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 901 TTCCGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960

Db 901 TTCCGGAGCTGAACAGCGCCACCCAGACTTCTGGAGGTGAGCTGGGACATCCCCAC 960
 Qy 961 CCCTCGGCGCTGAAGAGAGAGAGCGTGAACGCTGTCGAGCGTGGCGAGCGCTTACTTC 1020
 Db 961 CCCGCGGCGCTGAAGAGAGAGAGCGTGAACGCTGTCGAGCGTGGCGAGCGCTTACTTC 1020
 Qy 1021 AGCTGCCCCCTGAGCAGAGACTTCCGCAAGTACACCGCTTCACTCCCAAGATCAAC 1080
 Db 1021 AGCTGCCCCCTGAGCAGAGACTTCCGCAAGTACACCGCTTCACTCCCAAGATCAAC 1080
 Qy 1081 AACGAGACCCCGGCAATCCGCTACAGTCAACGCTGCGCCCAAGGAGTGAAGGGCAG 1140
 Db 1081 AACGAGACCCCGGCAATCCGCTACAGTCAACGCTGCGCCCAAGGAGTGAAGGGCAG 1140
 Qy 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTGAGGCCCTTCCGCGCCGCAACCCC 1200
 Db 1141 CCCAGCATTTCCAGAGCAGCATGACCAAGATCTGAGGCCCTTCCGCGCCGCAACCCC 1200
 Qy 1201 GAGATGTTGATTTACCA-----GGCCCCCTGTACTGTGGGCGAGCACTGAGATCGGC 1254
 Db 1201 GAGATGTTGATTTACCA-----GGCCCCCTGTACTGTGGGCGAGCACTGAGATCGGC 1260
 Qy 1255 CAGCACCGGCGCAAGTCCAGAGCTGCGGAGACACTGCTGCGTGGGCTTCAACACC 1314
 Db 1261 CAGCACCGGCGCAAGTCCAGAGCTGCGGAGACACTGCTGCGTGGGCTTCAACACC 1320
 Qy 1315 CCCGACAGAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1368
 Db 1321 CCCGACAGAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 Qy 1369 GACAGCTGAGCCTGCAAGCTGCAAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1428
 Db 1381 GACAGCTGAGCCTGCAAGCTGCAAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 Qy 1429 ATCCAGAGCTGTGGGCAAGCTGCAAGCTGTGGGCAAGCTGTACCCCGGCAACAGAGT 1488
 Db 1441 ATCCAGAGCTGTGGGCAAGCTGCAAGCTGTGGGCAAGCTGTACCCCGGCAACAGAGT 1500
 Qy 1489 CGCAGCTGTGCAAGCTGTGCGCGGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1548
 Db 1501 CGCAGCTGTGCAAGCTGTGCGCGGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
 Qy 1549 GAGAGAGCGGAGCTGAGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
 Db 1561 GAGAGAGCGGAGCTGAGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 Qy 1609 GTGTACTAGAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1668
 Db 1621 GTGTACTAGAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 Qy 1669 TGGAGCTTACAGATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
 Db 1681 TGGAGCTTACAGATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 Qy 1729 ATGCGCACCGCCCAACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1788
 Db 1741 ATGCGCACCGCCCAACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 Qy 1789 ATGAGAGAGATCTGTGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848
 Db 1801 ATGAGAGAGATCTGTGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 Qy 1849 ACCTGGAGAGAGCTGTGAGACCACTACTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1908
 Db 1861 ACCTGGAGAGAGCTGTGAGACCACTACTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
 Qy 1909 GTGAGACACCCCGCTGTGAGAGCTGTGATCCAGCTGAGAGAGAGAGAGAGAGAGAGAG 1968
 Db 1921 GTGAGACACCCCGCTGTGAGAGCTGTGATCCAGCTGAGAGAGAGAGAGAGAGAGAGAG 1980
 Qy 1969 GCCGAGAGCTTTCAAGTGAAGCGGCGGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAGAG 2028

Db 1981 GCCGAGAGCTTCTAGTGAAGCGGCGGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAG 2040
 Qy 2029 TACGTGACCGAGCCGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
 Db 2041 TACGTGACCGAGCCGGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
 Qy 2089 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2148
 Db 2101 ACCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 Qy 2149 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2208
 Db 2161 ACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 Qy 2209 CTGTGAGACCAAGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2268
 Db 2221 CTGTGAGACCAAGATCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
 Qy 2269 CCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2328
 Db 2281 CCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
 Qy 2329 CGCAGAGCTGTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2388
 Db 2341 CGCAGAGCTGTCTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
 Qy 2389 GACCTGTACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2448
 Db 2401 GACCTGTACGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
 Qy 2449 GGTGAATTC 2457
 Db 2461 GGTGAATTC 2469

RESULT 4
 BD263705
 LOCUS
 DEFINITION
 Improved expression of HIV polyproteins and production of
 virus-like particles.
 ACCESSION
 BD263705
 VERSION
 BD263705.1 GI:33073473
 KEYWORDS
 JP 2002533124-A/72.
 SOURCE
 synthetic construct
 ORGANISM
 other sequences; artificial sequences.
 REFERENCE
 1 (bases 1 to 2300)
 AUTHORS
 Barnett, S., Megede, J. Z., Sriwastava, I., Lian, Y., Hartog, K., Liu, H.,
 Greer, C., Selby, M. and Walker, C.
 TITLE
 Improved expression of HIV polyproteins and production of
 virus-like particles
 JOURNAL
 Patent: JP 2002533124-A 72 08-OCT-2002;
 CHIRON CORP
 COMMENT
 OS Artificial Sequence
 PM JP 2002533124-A/72
 PD 08-OCT-2002
 PE 30-DEC-1999 JP 2000591193
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI
 SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIWASTAVA, YING LIAN, PI
 KARIN HARTOG,
 PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC
 C12N5/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
 C12N5/10,
 PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
 Description of Artificial Sequence: Pst(-). protomed, Rtopc, YMMW FH
 Key Location/Qualifiers
 FT source 1..2300
 FT Location/Qualifiers
 FEATURES
 source
 1..2300
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN

Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best Local Similarity 93.6%; Pred. No. 6.3e-167;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

QY 170 GCGGCAAGAGGCGCCACAGATGAAGAGCTGCAACGAGCGCCAGGCCAATTCTTCCGG 229
DB 1 GCGGCGCGGAAGAGACCCAAATGAATGTCATGAGAGACAGGCTAATTTCTTCCGG 60
QY 230 AGGACTTGCCCTTCCCCCGAGGCAAGGCCCGGAGTTCCCGAGGCAAGAACCGCGCA 289
DB 61 AGGACTTGCCCTTCTGCAAGGCGCAAGGCCCGGAGTTCAAGCAAGCAAGACCGCGCA 120
QY 290 ACAAGCCCAACGAGCGCGAGCTGCAAGTGGCGGCG-----ACAACCCCGCAAGAG 343
DB 121 ACAAGCCCAACGAGCGCGAGCTGCAAGTGGCGGCGCGAGAACAAACAGCTGAAGAG 180
QY 344 CGGCGCGCCAGAGCGCAAGGCAACCTG-----AATTCCCGAGATCAACCTGTGGCAG 397
DB 181 CGGCGCGCGACCGCGCAAGGCAACGTAAGCTTCACTTCCCGAGATCAACCTGTGGCAG 240
QY 398 GCCCTCTGTGATCAAGTGGCGCGCAAGTCAAGAGGCCCTGTGCAACCGCG 457
DB 241 GCCCTCTGTGATCAAGTGGCGCGCAAGTCAAGAGGCCCTGTGCAACCGCG 300
QY 458 CGGAGCAACCGTGTGAGAGATGAGCTGCCGCGCAAGTGAAGCCCAAGATGATCG 517
DB 301 CGGAGCAACCGTGTGAGAGATGAGCTGCCGCGCAAGTGAAGGCCCAAGATGATCG 360
QY 518 GCGGCAACGCGGCTTCAAGAGTGGCGCAAGTCAAGAGTCTGATCGAGATCTGCG 577
DB 361 GCGGCAACGCGGCTTCAAGAGTGGCGCAAGTCAAGAGTCTGATCGAGATCTGCG 420
QY 578 GCGAAGAGGCTATCGGCAACGCTGCTGATGCGCCCAACCCCGTGAACATCATCGCGCA 637
DB 421 GCGAAGAGGCTATCGGCAACGCTGCTGCTGATGCGCCCAACCCCGTGAACATCATCGCGCA 480
QY 638 ACATGCGGCGCGAGCTGAGCTGCAACCTGAACCTTCCCGATCAAGCCCATCGAGCGTGC 697
DB 481 ACATGCGGCGCGAGCTGAGCTGCAACCTTGAACCTTCCCGATCAAGCCCATCGAGCGTGC 540
QY 698 CGGTGAAGCTGAAGCCCGCAATGAGCGGCCCAAGGTGAAGCATGCGCCCTGACCGAG 757
DB 541 CGGTGAAGCTGAAGCCCGGAGTGAAGCGGCCCAAGGTGAAGCATGCGCCCTGACCGAG 600
QY 758 AGAAGATCAAGGCCCTGAGCGCATTTGCGAGAGATGGAAGAGAGGCCAAGATCAACA 817
DB 601 AGAAGATCAAGGCCCTGAGCGATCTGCAACGAGATGGAAGAGAGGCCAAGATCAACA 660
QY 818 AGATGCGGCGCGAGAACCCCTGAACAACCCCGTGTGGCATTCAGAGAGAGAGCAAGA 877
DB 661 AGATGCGGCGCGAGAACCCCTGAACAACCCCGTGTGGCATTCAGAGAGAGAGCAAGA 720
QY 878 CCAAGTGGCGCAAGCTGTGACTTCCGAGAGCTGAACAAGGCAACCGAGCTTCTGG 937
DB 721 CCAAGTGGCGCAAGCTGTGACTTCCGAGAGCTGAACAAGGCAACCGAGCTTCTGG 780
QY 938 AGGTGAGCTGGGACATCCCGCAACCCCGCGCGCTGAAGAGAGAGAGAGAGCTGACCGTGC 997
DB 781 AGGTGAGCTGGGACATCCCGCAACCCCGCGCGCTGAAGAGAGAGAGAGAGAGCTGACCGTGC 840
QY 998 TGAAGGTGGGCGACCGCTACTTCAAGGTGCCCTTGAAGAGAGCTTCCGAGAGTCAACG 1057
DB 841 TGAAGGTGGGCGACCGCTACTTCAAGGTGCCCTTGAAGAGAGCTTCCGAGAGTCAACG 900
QY 1058 CCTTCAACATCCCGAGCATCAACAAGAGACCCCGGCGATCCGCTACAGTCAACAGTGC 1117
DB 901 CCTTCAACATCCCGAGCATCAACAAGAGACCCCGGCGATCCGCTACAGTCAACAGTGC 960
QY 1118 TGCCTCAAGGCTGGAAGGCGAGCCCGAGCATTTCCAGAGAGAGATGACCAAGATCTGG 1177
DB 961 TGCCTCAAGGCTGGAAGGCGAGCCCGAGCATTTCCAGAGAGAGATGACCAAGATCTGG 1020

QY 1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTAACAGGCGCCCTGTACGTGGCA 1237
DB 1021 AGCCCTTCCGCGAGAGAGAACCCCGACATCGTATCTTACAGGCGCCCTGTACGTGGCA 1080
QY 1238 GCGACTTGAAGTCGGCGAGCAACCGCGCAAGATGAGAGAGCTGCGCAAGCACTGTGC 1297
DB 1081 GCGACTTGAAGTCGGCGAGCAACCGCGCAAGATGAGAGAGCTGCGCGAGCACTGTGC 1140
QY 1298 GCTGGGCGCTTCAACACCCCGCAAGAGAGCAACAGAGCGCCCTTCCGTGCCATCG 1357
DB 1141 GCTGGGCGCTTCAACACCCCGCAAGAGAGCAACAGAGAGCGCCCTTCCGTGCCATCG 1200
QY 1358 AGCTGCAACCCCGCAAGTGAACCGTGCAGGCCCATCGAGCTGCGCGAGAGAGAGAGTGA 1417
DB 1201 AGCTGCAACCCCGCAAGTGAACCGTGCAGGCCCATCATCTGTGCGAGAGAGAGAGTGA 1260
QY 1418 CGGTGAAGCAATCCAGAGAGCTGTGGCAAGCTGAACCTGGCCAGCGATCTACCCCG 1477
DB 1261 CGGTGAAGCAATCCAGAGAGCTGTGGCAAGCTGAACCTGGCCAGCGATCTACCGCG 1320
QY 1478 GCATCAAGTGGCGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCCCTGTACCGACATCG 1537
DB 1321 GCATCAAGTGAAGAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCCCTGTACCGAGTGA 1380
QY 1538 TGCCTCTGAACCGAGAGCGCGAGCTGGAAGCTGGCGAGAACCGCGAGATCTTGGCGAGC 1597
DB 1381 TGCCTCTGAACCGAGAGCGCGAGCTGGAAGCTGGCGAGAACCGCGAGATCTTGAAGAGC 1440
QY 1598 CGGTCAACGCGGTGTACTTACAGACCCCGCAAGAGCATTTGGTGGCGAGATTCAGAAAGCAG 1657
DB 1441 CGGTCAACGAGGTGTACTTACAGACCCCGCAAGAGCATTTGGTGGCGAGATTCAGAAAGCAG 1500
QY 1658 GCGAAGCAAGTGAACCTTACAGAGATCTTACAGAGAGCCCTTGAAGAACCTTGAAGACCGGCA 1717
DB 1501 GCGAAGGCGCAAGTGAACCTTACAGAGATCTTACAGAGAGCCCTTGAAGAACCTTGAAGACCGGCA 1560
QY 1718 AGTACGCGCAAGATGCGCACCGCGCAACCAACAGAGTGAAGAGCTGACCGAGCGTGC 1777
DB 1561 AGTACGCGCGATGCGCGCGCGCGCGCGCACCAACAGAGTGAAGAGCTGACCGAGCGTGC 1620
QY 1778 AGAAGATGCGCACATGAGAGCATGTGATCTGGGCGAAGACCCCGAAGTTCCGCTGCCCA 1837
DB 1621 AGAAGATGAGCACCGAGAGCATGTGATCTGGGCGAAGATTCGCAAGTTCAAGCTGCCCA 1680
QY 1838 TTCAGAGAGAGACTTGGAGAGACTTGTGACCGACTTACGTGAGAGCGCATCTGGAATCCCG 1897
DB 1681 TTCAGAGAGAGACTTGGAGAGCGCTGTGATGAGAGTACGTGAGAGCGCATCTGGAATCCCG 1740
QY 1898 AGTGGAGTGTGTGAACACCCCGCGCTGTGAGAGCTGTGTGAACAGCTGAGAGAGAGC 1957
DB 1741 AGTGGAGTGTGTGAACACCCCGCGCTGTGTGAGAGCTGTGTGAACAGCTGAGAGAGAGC 1800
QY 1958 CCATCATCGCGCGCGAGACTTCTTACGTGAGAGCGCGCGCGCAACCGCGAGACCAAGATCG 2017
DB 1801 CCATGTGGGCGCGAGACTTCTTACGTGAGAGCGCGCGCGCAACCGCGAGACCAAGCTGC 1860
QY 2018 GCAAGCGCGCTTACGTGACCGACCGGCGCGCGCAAGAGTCTGTAGGCTGACCGAGACCA 2077
DB 1861 GCAAGCGCGCTTACGTGACCGACCGGCGCGCGCAAGAGTGTGTGAGTCTGCGCAACCA 1920
QY 2078 CCAACAGAGAGCGAGCTGCAAGGCGCATTCAGCTGAGCGCCCTGAGAGAGCGCGCAAGAG 2137
DB 1921 CCAACAGAGAGCGAGCTGCAAGGCGCATTCAGCTGAGAGCGCGCCCTGAGAGAG 1980
QY 2138 TGAACATGTGACCGAGAGCGAGTACGCTGTGGCATCATCCAGGCGCGAGCGCGCAAGA 2197
DB 1981 TGAACATGTGACCGAGAGCGAGTACGCTGTGGCATCATCCAGGCGCGAGCGCGCAAGA 2040
QY 2198 GCGAGAGGAGCTGTGTAACCAAGATCATGAGAGCTGTATCAAGAGAGAGAGTGTACC 2257
DB 2041 GCGAGAGGAGCTGTGTAACCAAGATCATGAGAGCTGTATCAAGAGAGAGAGTGTACC 2100

QY 2258 TGAAGCTGGTGGCCGCCCAAGAGGAGTGGGCGAAGAGAGATCCAGAGCTGTGTA 2317
 DB 2101 TGGCTGGGGTCCCGCCCAAGAGGAGTGGGCGAAGAGAGTGGAGCAAGCTGTGTA 2160
 QY 2318 GCAAGGGCATCCGCAAGAGTGTCTTCTGAGACGGCATCGATGCGGATCGTATAC 2377
 DB 2161 GGGCGGCAATCCGCAAGAGTGTCTTCTGAGACGGCATCGATGCGGATCGTATAC 2220
 QY 2378 AGTACATGAGCACTGTACGTGGGAGCGCGGCTTGAATTCATTAAGCTTCCG 2437
 DB 2221 AGTACATGAGCACTGTACGTGGGAGCGCGGCTTGAATTCATTAAGCTTCCG 2280
 QY 2438 GGGCTAGCACCGGTGAATTC 2457
 DB 2281 GGGCTAGCACCGGTGAATTC 2300
 RESULT 5
 LOCUS CQ870575 2300 bp DNA linear PART 13-SEP-2004
 DEFINITION Sequence 83 from Patent EP1433851.
 ACCESSION CQ870575
 VERSION CQ870575.1 GI:52000091
 KEYWORDS
 SOURCE Synthetic construct
 ORGANISM synthetic construct
 other sequences; artificial sequences.
 REFERENCE 1
 AUTHORS Barnett, S., Zumedege, J., Srivastava, I., Lian, Y., Hartog, K.,
 Liu, H., Greer, C., Selby, M. and Walker, C.
 TITLE Improved expression of HIV polypeptides and production of
 JOURNAL virus-like particles
 Patent: EP 1433851-A 83 30-JUN-2004;
 CHIRON CORPORATION (US)
 FEATURES
 source 1..2300
 Location/Qualifiers
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Description of Artificial Sequence:
 FS(-).protomod.RTocp.YMMM"
 ORIGIN
 Query Match 83.0%; Score 2040; DB 6; Length 2300;
 Best Local Similarity 93.6%; Pred. No. 6.3e-167;
 Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
 QY 170 GCGGCAAGAGGGCCACGATGAGAGTGCACCGAGCGCCCACTTCTTCCGCG 229
 DB 1 GCGGCGCGAAGAGCACTAATGAAGATTGCACTGAGAGACAGGCTTAATTTCTTCCGCG 60
 QY 230 AGGACCTGGCTTCCCGCAAGGAGGCGCGAGTTCGCCAGAGACAGACCGGCGCA 289
 DB 61 AGGACCTGGCTTCTGCAAGGAGGCGCGAGTTCGCCAGAGACAGACCGGCGCA 120
 QY 290 ACAAGCCCAACCAAGCGGAGCTGAGTGCAGCGG-----ACAAGCCCGGAGAGG 343
 DB 121 ACAAGCCCAACCGCGGAGCTGAGTGCAGCGGCGGCGAGAACAGAGCTGAGCGAG 180
 QY 344 CCGGCGCGAGCGCGAGAGGCACTCTG-----AACTTCCCGGAGATCACTTGTGCAAC 397
 DB 181 CCGGCGCGAGCGCGAGAGGCACTCTG-----AACTTCCCGGAGATCACTTGTGCAAC 240
 QY 398 GCGCCCTGAGATCAAGTGGGCGGCGAGATCAAGAGGCGCTGCTGAGACCGGCG 457
 DB 241 GCGCCCTGAGATCAAGTGGGCGGCGAGATCAAGAGGCGCTGCTGAGACCGGCG 300
 QY 458 CCGAGCAACCGTGTCTGAGAGAGTGAAGCTTCCCGGCAAGTGAAGCCCAAGATATCG 517
 DB 301 CCGAGCAACCGTGTCTGAGAGAGTGAAGCTTCCCGGCAAGTGAAGCCCAAGATATCG 360
 QY 518 GCGGCAATGGCGGCTTATCAAGTGGCGAGTACAGACAGATCTGATGAGATTCGCG 577
 DB 361 GCGGCAATGGCGGCTTATCAAGTGGCGAGTACAGACAGATCTGATGAGATTCGCG 420
 QY 578 GCAAGAGGCGCATCGGCAACGTGTGATCGGCGCCCAACCCCGGTGAACATTCGCGCGCA 637
 DB 421 GCAAGAGGCGCATCGGCAACGTGTGATCGGCGCCCAACCCCGGTGAACATTCGCGCGCA 480
 QY 638 ACATCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATGAGCCCATCGAGACCGTGC 697
 DB 481 ACATCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATGAGCCCATCGAGACCGTGC 540
 QY 698 CCGTGAAGCTGAAGCCCGGCAATGAGCGGCGCCCAAGTGAAGAGTGGCCCTGACCGAG 757
 DB 541 CCGTGAAGCTGAAGCCCGGCAATGAGCGGCGCCCAAGTGAAGAGTGGCCCTGACCGAG 600
 QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGAGATGAGAGAGGCGCAAGATCAACA 817
 DB 601 AGAAGATCAAGGCGCTGAGATCTGACCGCATGAGAGAGAGGCGCAAGATCAACA 660
 QY 818 AGATGGGCGCGAGAACCCCTTACAACACCCCGGTGGCCATCAAGAGAGAGCAGCA 877
 DB 661 AGATGGGCGCGAGAACCCCTTACAACACCCCGGTGGCCATCAAGAGAGAGCAGCA 720
 QY 878 CCAAGTGGCGCAAGCTGTGATCTTCCGAGAGCTGAACAAGCGCAACCGAGACTTCTGCG 937
 DB 721 CCAAGTGGCGCAAGCTGTGATCTTCCGAGAGCTGAACAAGCGCAACCGAGACTTCTGCG 780
 QY 938 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGTGCCTGTC 997
 DB 781 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGTGCCTGTC 840
 QY 998 TGAAGTGGGCGAACCCTTACCTTCAAGTGGCGCTGAGAGAGAGACTTCCGAGATCAACCG 1057
 DB 841 TGAAGTGGGCGAACCCTTACCTTCAAGTGGCGCTGAGAGAGAGACTTCCGAGATCAACCG 900
 QY 1058 CTTTCAACATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
 DB 901 CTTTCAACATCCCGAGCATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
 QY 1118 TGCAGGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1177
 DB 961 TGCAGGAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1178 AGCCCTTCCGCGCGCGGCAACCCCGAGATGTGATCTACAGAGAGAGAGAGAGAGAGAG 1237
 DB 1021 AGCCCTTCCGCGCGGCAACCCCGAGATGTGATCTACAGAGAGAGAGAGAGAGAGAG 1080
 QY 1238 GGAAGCTGGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1297
 DB 1081 GGAAGCTGGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 QY 1298 GCTGGGCTTCAACACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
 DB 1141 GCTGGGCTTCAACACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1358 AGCTGACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
 DB 1201 AGCTGACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1418 CCGTGAAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
 DB 1261 CCGTGAAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1478 GCATCAAGTGGCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
 DB 1321 GCATCAAGTGGCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
 QY 1538 TGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
 DB 1381 TGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1598 CCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
 DB 1441 CCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500

QY 1658 GCCACGACGAGTGAAGCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAAGACCGGCA 1717
 DB 1501 GCCAGGAGCCAGTGAAGCTTACAGATCTTACAGAGAGCCCTTCAAGAACTTGAAGACCGGCA 1560
 QY 1718 AGTACGCCAAGATGCGACCGGCTCAACCAAGACGTGAAGAGCTGAACCGAGCCGTGC 1777
 DB 1561 AGTACGCCCGACGCGCGCGGCCCAACCAAGACGTGAAGAGCTGAACCGAGCCGTGC 1620
 QY 1778 AGAAGATCGCCATGAGAGCACTGTGATCTTGGGGGCAAGACCCCAAGTTCCGCTGCCCA 1837
 DB 1621 AGAAGATGAGCAACCGAGACGATGTGATCTGAGGCAAGATCCCAAGTTCAAGCTGCCCA 1680
 QY 1838 TCCAGAGAGAGACTTGGAGAGACTTGTGTGACCGACTTGTGACCGACTTGTGACCGACTTGTGAC 1897
 DB 1681 TCCAGAGAGAGACTTGGAGAGACTTGTGTGACCGACTTGTGACCGACTTGTGACCGACTTGTGAC 1740
 QY 1898 AGTGGAGTTCTGTGAACACCCCTGTGTGAGCTGTGTGACCGACTTGTGACCGACTTGTGACCG 1957
 DB 1741 AGTGGAGTTCTGTGAACACCCCTGTGTGAGCTGTGTGACCGACTTGTGACCGACTTGTGACCG 1800
 QY 1958 CCATCATGCGGCGCGAGACCTTCTAGTGTGAGCGCGCGCGCAACCGCGAGACCAAGATCG 2017
 DB 1801 CCATGCTGCGCGCGAGACCTTCTAGTGTGAGCGCGCGCGCAACCGCGAGACCAAGCTGCG 1860
 QY 2018 GCAAGGCGCGCTACGTGACCGACCGGCGCGCGCAAGAGATCGTGAAGCTTGAACCGAGACCA 2077
 DB 1861 GCAAGGCGCGCTACGTGACCGACCGGCGCGCGCAAGAGATCGTGAAGCTTGAACCGAGACCA 1920
 QY 2078 CCAACCGAGAGACCGAGCTGAGAGCCATCACTGAGCTGAGAGACCGGCGAGCGAG 2137
 DB 1921 CCAACCGAGAGACCGAGCTGAGAGCCATCACTGAGCTGAGAGACCGGCGAGCGAGCTGAGAG 1980
 QY 2138 TGAATCATGCTGACCGACAGCCGATGACCGCTGTGAGATCTTCAAGGCGCGCGAGACCA 2197
 DB 1981 TGAATCATGCTGACCGACAGCCGATGACCGCTGTGAGATCTTCAAGGCGCGCGAGACCA 2040
 QY 2198 GCGAGAGCGAGCTGTGAAACGATCATGAGCAGCTGATCAAGAGAGAGAGTGTACC 2257
 DB 2041 GCGAGAGCGAGCTGTGAAACGATCATGAGCAGCTGATCAAGAGAGAGAGTGTACC 2100
 QY 2258 TGAAGTGTGAGTCCGCGCCCAAGAGGATGTGCGGCAACGAGACGATCGAACAGCTGTGA 2317
 DB 2101 TGAAGTGTGAGTCCGCGCCCAAGAGGATGTGCGGCAACGAGACGATCGAACAGCTGTGA 2160
 QY 2318 GGAAGGCGATCCGCAAGGCTGTCTTGGAGAGCGATCGATGCGCGCATCTGTGATCTTACC 2377
 DB 2161 GCGCGCGCATCCGCAAGGCTGTCTTGGAGAGCGATCGATGCGCGCATCTGTGATCTTACC 2220
 QY 2378 AGTACATGAGACGCTGTAGTGTGAGCGAGCGCGCGCTTGAAGATCGATTAAAGCTTCCCG 2437
 DB 2221 AGTACATGAGACGCTGTAGTGTGAGCGAGCGCGCGCGCTTGAAGATCGATTAAAGCTTCCCG 2280
 QY 2438 GGGCTAGACCGCGGTGAATTC 2457
 DB 2281 GGGCTAGACCGCGGTGAATTC 2300

RESULT 6
 AR373388 2300 bp DNA linear PAT 18-DEC-2003
 LOCUS AR373388
 DEFINITION Sequence 83 from patent US 6602705.
 ACCESSION AR373388
 VERSION AR373388.1 GI:40075491
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCES 1 (bases 1 to 2300)
 AUTHORS Barnett,S.W., Megede,J.J., Greer,C. and Selby,M.
 TITLE Expression of HIV polypeptides and production of virus-like particles
 JOURNAL Patent: US 6602705-A 83 05-AUG-2003;

FEATURES Chiron Corporation; Emeryville, CA
 source Location/Qualifiers
 1..2300
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 83.0%; Score 2040; DB 6; Length 2300;
 Best Local Similarity 93.6%; Pred. No. 6.3e-167;
 Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

QY 170 GCGGCAAGAGAGGCCCAAGATGAAGACTGACCGAGCGCAAGCTTCTTCCGCG 229
 DB 1 GCGGCGGGAAGAGACCAAAATGAAGATTCAGTGAAGAGACAGGCTAATTTCTTCCGCG 60
 QY 220 AGGACTTGGCTTCCCCCAAGGCGCAAGGCGCGAGTTCCCCAGAGAGAGAAACCGCGCA 289
 DB 61 AGGACTTGGCTTCTTGAAGGCGCAAGGCGCGAGTTCAAGAGAGAGAGACCGCGCA 120
 QY 290 ACAAGCCCAACGCGCAGCTGCAAGTGCAGCGGCG-----ACAACCCCGCAGCGAG 343
 DB 121 ACAAGCCCAACGCGCAGCTGCAAGTGCAGAGTGTGGGCGCGAGAGAACAGCTTGAAGCGAG 180
 QY 344 CCGGCGCGAGCGCGAGGCGACCTTG-----AATTCCCGCAAGTCAACCTTGTGCGAGC 397
 DB 181 CCGGCGCGAGCGCGAGGCGACCTTGAGCTGTGGGCGCGAGAGAACAGCTTGAAGCGAGC 240
 QY 398 GCCCTCTGTGAGATCAAGAGTGGCGCGCGAGATCAAGAGAGCCCTGTGACACCGCGC 457
 DB 241 GCCCTCTGTGATCAATCAAGATTCGGCGCGCGAGCTTCAGAGAGAGCGCTGTGACACCGCGC 300
 QY 458 CCGAGCAACCGGTGTGAGAGAGATGAGCTGCGCGCGAGATGAGAGCCCAAGATGATCG 517
 DB 301 CCGAGCAACCGGTGTGAGAGAGATGAGCTGCGCGCGAGATGAGAGCCCAAGATGATCG 360
 QY 518 GCGGATGCGGCGCTTCAATCAAGGTGCGCGAGTACAGACAGATCTTGAATCGAGATCTGCG 577
 DB 361 GCGGATGCGGCGCTTCAATCAAGGTGCGCGAGTACAGACAGATCTTGAATCGAGATCTGCG 420
 QY 578 GCAAGAGGCGATGCGACCGGTGTGATCGCGCGCGCAACCCCGTGAACATCATCGCGCGCA 637
 DB 421 GCAAGAGGCGATGCGACCGGTGTGATCGCGCGCGCAACCCCGTGAACATCATCGCGCGCA 480
 QY 638 ACATGCTGACCGAGCTGTGACCGCTTGAACCTTCCCATGAGCGCCCATTCGAGACCGTGC 697
 DB 481 ACATGCTGACCGAGCTGTGACCGCTTGAACCTTCCCATGAGCGCCCATTCGAGACCGTGC 540
 QY 698 CCGTGAAGCTGAAGCGCGCGCATGGAACCGCGCCCAAGGTGAAGCAAGTGCCTTGAACCGAGG 757
 DB 541 CCGTGAAGCTGAAGCGCGCGCATGGAACCGCGCCCAAGGTGAAGCAAGTGCCTTGAACCGAGG 600
 QY 758 AGAAGATCAAGGCGCTTGAACCGCGCATGTGCGAGAGAGATGAAGAGAGGCGCAAGATCAACA 817
 DB 601 AGAAGATCAAGGCGCTTGTGTGAGATGTGCAACGAGATGAAGAGAGGCGCAAGATCAACA 660
 QY 818 AGATGCGCGCGAGAGACCGCTTACAAACCGCGGTGTGCGCATCAAGAGAGAGAGACAGA 877
 DB 661 AGATGCGCGCGAGAGACCGCTTACAAACCGCGGTGTGCGCATCAAGAGAGAGAGAGACAGA 720
 QY 878 CCAAGTGAACGAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCAACCGAGACTTCTGAGG 937
 DB 721 CCAAGTGAACGAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCAACCGAGACTTCTGAGG 780
 QY 938 AGGTGCACTGGGCGATCCCCCAACCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 997
 DB 781 AGGTGCACTGGGCGATCCCCCAACCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 840
 QY 998 TGAAGTGGGAGAGCGCTTCAAGCGTGGCGCGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
 DB 841 TGAAGTGGGAGAGCGCTTCAAGCGTGGCGCGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 1058 CTTTCAACCATCCGAGATCAACAGAGACCGCGCGCATCGCTACAGAGTCAACAGTGC 1117

Db 901 CCTTCAACCATCCCAAGCATCAACAAGAGAACCCCCGGCATCCGGTACAGTACAAAGTGC 960
 QY 1118 TCCCCAGGGCTGTGAAGGGGACGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1177
 Db 961 TCCCCAGGGCTGTGAAGGGGACGCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGG 1020
 QY 1178 AGCCCTCCGGCCCGCCAAACCCCGAGATGTGTATCTAACAGGCCCCCTGTATCTGGGGA 1237
 Db 1021 AGCCCTCCGGCAAGCAAAACCCCGACATGTATCTAACAGGCCCCCTGTATCTGGGGA 1080
 QY 1238 GCGACCTGTGAATCGGCACGACACCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGC 1297
 Db 1081 GCGACCTGTGAATCGGCACGACACCGCACTGAATCGAGAGCTGCGCCAGCACTGCTGC 1140
 QY 1298 GCTGGGGCTTCAACACCCCGCAAGAGACACCAAGAGAGCCCTTCTGCTCCATCG 1357
 Db 1141 GCTGGGGCTTCAACACCCCGCAAGAGACACCAAGAGAGCCCTTCTGCTCCATCG 1200
 QY 1358 AGCTGACCCCGCAAGTGAACCTGTGCAAGCCCATCGAGTGGCCCGAGAGAGAGCTGGA 1417
 Db 1201 AGCTGACCCCGCAAGTGAACCTGTGCAAGCCCATCGAGTGGCCCGAGAGAGAGCTGGA 1260
 QY 1418 CCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAATGGGCGACGACGATCTACCCG 1477
 Db 1261 CCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAATGGGCGACGACGATCTACCCG 1320
 QY 1478 GCATCAAGGTGCGCAGCTGTGCAAGCTGTGCGCGCGCCAGAGCCCTGACCGACATCG 1537
 Db 1321 GCATCAAGGTGCAAGCAGCTGTGCAAGCTGTGCGCGCGCCAGAGCCCTGACCGAGTGA 1380
 QY 1538 TGGCCCTTGACCGAGAGGCGCGAGCTGTGCGCGCGCAAGACCGCGAGATCTGCGCGAC 1597
 Db 1381 TGGCCCTTGACCGAGAGGCGCGAGCTGTGCGCGCGCAAGACCGCGAGATCTGGAAGAC 1440
 QY 1598 CCGTGACCGCGGTGATCTAGACCCAGCAAGGACCTGTGGCGCGAGATCCAGAACGAG 1657
 Db 1441 CCGTGACCGAGGTGATCTAGACCCAGCAAGGACCTGTGGCGCGAGATCCAGAACGAG 1500
 QY 1658 GCGCAGCAGCATGTGACCTTACAGANTCTACAGAGAGCCCTTCAAGAACCTGAAGACGGCA 1717
 Db 1501 GCGCAGGCGCAGTGAACCTTACAGANTCTTACAGAGAGCCCTTCAAGAACCTGAAGACGGCA 1560
 QY 1718 AGTACGCCAAGATGCGGACCGCCCAACAACAAGCTGAAGCAGCTGACCGAGGCGGTGC 1777
 Db 1561 AGTACGCCCGCATGCGCGCGCCCAACAACAAGCTGAAGCAGCTGACCGAGGCGGTGC 1620
 QY 1778 AGAAGATGCGCAGTGAAGAGATGTGTATCTGGGCGCAAGACCCCAAGTTCCGCTGCCA 1837
 Db 1621 AGAAGGTGAGACCGAGAGCATGTATCTGGGCGCAAGATCCCAAGTTCCAGCTGCCA 1680
 QY 1838 TCCAGAGAGAGACCTGTGAGACCTGTGTGACCGACTACTGTGCGAGCAGCTGTGATCCCG 1897
 Db 1681 TCCAGAGAGAGACCTGTGAGAGCTGTGTGATGTGAATGACTGTGCGACACTGTATCCCG 1740
 QY 1898 AGTGGAGTTCTGTGAACACCCCTCCCTGTGTGAAGCTGTGTGACAGCTGTGAGAGAGAG 1957
 Db 1741 AGTGGAGTTCTGTGAACACCCCTCCCTGTGTGAAGCTGTGTGACAGCTGTGAGAGAGAG 1800
 QY 1958 CCAATCATGCGCGCGAGACCTTCTAAGTGAACGCGCCCGCAACCGCGAGAACCAAGATCG 2017
 Db 1801 CCAATCTGTGCGCGCGAGACCTTCTAAGTGAACGCGCGCGCAACCGCGAGAACCAAGCTG 1860
 QY 2018 GCAAGCGCGCTAAGTGAACCGAGCGCGCGAGAGAGATCTGTAGCCTGTGACCGAGACCA 2077
 Db 1861 GCAAGCGCGCTAAGTGAACCGAGCGCGCGAGAGAGATGTGTAGATGTGCGAGACCA 1920
 QY 2078 CCAACCAAGAGACCGAGCTGCAAGGCGCATCAAGCTGTGCTGTGACAGACGCGGCGAGAG 2137
 Db 1921 CCAACCAAGAGACCGAGCTGCAAGGCGCATCAAGCTGTGCTGTGACAGACGCGGCTGTGAG 1980
 QY 2138 TGAACATGTGTGAACGACAGCAAGTGAACGCTGTGGCGCATCAAGGCGCGAGCGCAAGAA 2197
 Db 1981 TGAACATGTGTGAACGACAGCAAGTGAACGCTGTGGCGCATCAAGGCGCGAGCGCAAGAA 2040

QY 2198 GCGAGAGGAGCTGTGTGAACAGAGATCATGAGCAGCTGATCAAGAGAGAGAGGTGTAAC 2257
 Db 2041 GCGAGAGGAGCTGTGTGAACAGAGATCATGAGCAGCTGATCAAGAGAGAGAGGTGTAAC 2100
 QY 2258 TGAAGTGGGTGCGCGCCCAAGAGGAGCATCGCGGCAAGAGAGATGCAAGAGCTGTGTA 2317
 Db 2101 TGAAGTGGGTGCGCGCCCAAGAGGAGCATCGCGGCAAGAGAGATGCAAGAGCTGTGTA 2160
 QY 2318 GCAAGGAGATCGCAAGGTGCTGTCTGTGAACGCGCATGTATGCGGCGCATCTGTAC 2377
 Db 2161 GCGCGGAGATCGCAAGGTGCTGTCTGTGAACGCGCATGTATGCGGCGCATCTGTAC 2220
 QY 2378 AGTACATGAGAGACCTGTATCGTGGGCAAGCGCGCGCTAGAGATGCAATTAAAGCTTCCG 2437
 Db 2221 AGTACATGAGAGACCTGTATCGTGGGCAAGCGCGCGCTAGAGATGCAATTAAAGCTTCCG 2280
 QY 2438 GGGCTAGCACCGGTGAATTC 2457
 Db 2281 GGGCTAGCACCGGTGAATTC 2300

RESULT 7
 BD263704
 LOCUS
 DEFINITION Improved expression of HIV polypeptides and production of
 virus-like particles.
 ACCESSION BD263704
 VERSION BD263704.1 GI:33073472
 KEYWORDS JP 200253124-A/71.
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 OS Artificial Sequence
 PN JP 200253124-A/71
 PD 08-OCT-2002 JP 2000591193
 PF 30-DEC-1999 US 60/114495, 01-DEC-1999 US 60/168471 PI
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI
 SUSAN BARNETT, JAN ZUR MEGEBDE, INDRESH SRIVASTAVA, YING LIAN, PI
 KARIN HARTOG,
 PI HONG LIU CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC
 C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
 C12N5/10,
 PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
 Description of Artificial Sequence: FS(-). protomod. Rtopt. YM FH Key
 FT source 1. .2306
 Location/Qualifiers
 1. .2306
 /organism="Artificial Sequence"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN
 Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
 Best Local Similarity 93.2%; Pred. No. 3.9e-165;
 Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGGCCACCAAGATGAAGATGTGACCGAGCGCCAGGCTTCTTCCGGG 229
 Db 1 GCGGCGCGCACAAGAGACCAAAATGAAGATTGCACTGAAGAGACGCTAATTTCTTCCGG 60
 QY 230 AGGACCTGCGCTTCCCGCCAGAGGCGAGCGCGGAGATTCCCGAGCGAGCAAGACCGGCGCA 289
 Db 61 AGGACCTGCGCTTCTGCGAGGCGAGCGCGCGAGATTCAAGCGAGCGAGACCGCGCGCA 120

OY	290	CAAGCCCAACAAGCCGAGAGCTGACGTCGCGCGG-----ACACCCCGGACGAG	343
Db	121	ACAGCCCAACCGCGCGAGCTGACAGGTGTGGGGGCGAGAACAAACAGCTTGAGCGAG	180
OY	344	CCGGCGCCGACGCGCCAGGGCAACCTCG-----AACCTCCCGAGATCAACCTGTGGAGC	397
Db	181	CCGGGCGCCGACCGCCAGGGCAACCGTGAACCTTCACTTCCCGAGATCAACCTGTGGAGC	240
OY	398	GGCCCTGTGTAGCATCAAGTGGCGGGCCAGATCAAGAGAGCCCTGTGTGACACCGGCG	457
Db	241	GGCCCTGTGTATCAATCAAGATTCGGCGGCGACAGCTCAAGAGAGCGCTGTGTCAACCGGCG	300
OY	458	CCGACGACCAACGTCGTGAGAGAGATGACCTGGCCCGGCAAGTGGAGAGCCCAAGATGATCG	517
Db	301	CCGACGACCAACGTCGTGAGAGAGATGAACTGGCCCGGCAAGTGGAGAGCCCAAGATGATCG	360
OY	518	GGCGATCGCGGCTTCATCAAGGTGCGCCAGTACGACCAATCTGTATCGATCTGCG	577
Db	361	GGCGGATTCGGGGGCTTCATCAAGGTGCGCGCAGTACGACCAATCTGTGTATGATCTGCG	420
OY	578	GCAAGAGAGCCATCGGCAACCGTGTGATTCGGCCCAACCCCGGTAACTCATTCGGCGCA	637
Db	421	GCACAAAGGCATCGGCAACCGTGTGATTCGGCCCAACCCCGGTAACTCATTCGGCGCA	480
OY	638	ACATCTGACCCAGCTGGGCTGCAACCTGAACTTCCCATGACGCCCATCGAGACCGTGC	697
Db	481	ACCTGTACCCAGATTCGGCTGCAACCTGAACTTCCCATGACGCCCATCGAGACCGTGC	540
OY	698	CCGTGAACTGAAAGCCCGGCATTCGACCGGCCCAAGTGAAGCATGTGGCCCTTCACCGAG	757
Db	541	CCGTGAACTGAAAGCCCGGAGATTCGACCGGCCCAAGTGAAGCATGTGGCCCTTCACCGAG	600
OY	758	AGAAATCAAGGCGCTTGAACCGGCATCTCGAGAGATGAGAAAGAGGGCAAGATCAACA	817
Db	601	AGAAATCAAGGCGCTTGTGTGAGATCTGCAACGAGATGAGAGAGGGCAAGATCAACA	660
OY	818	AGATCGGCGCCGAGAACCTCTTACAAACACCCCGTGTTCGCATCAAGAAAGAGACAGCA	877
Db	661	AGATCGGCGCCGAGAACCTCTTACAAACACCCCGTGTTCGCATCAAGAAAGAGACAGCA	720
OY	878	CCAATGGCGGCAAGTCGTGTGATCTTCGCGGAGCTGAAACAAACGCAACCGAGACTTTCGGG	937
Db	721	CCAATGGCGGCAAGTCGTGTGATCTTCGCGGAGCTGAAACAAACGCAACCGAGACTTTCGGG	780
OY	938	AGGTGCACTGGGCAATCCGCCACCCCGCGGCTTGAAAGAAAGAGAGCTGACCGTGC	997
Db	781	AGGTGCACTGGGCAATCCGCCACCCCGCGCTTGAAAGAAAGAGAGCTGACCGTGC	840
OY	998	TGGAAGTGGGGGAGACGCTTACTTCAAGCGTGGCCCTGGAAGAGACTTCCGACAGTACACCG	1057
Db	841	TGGAAGTGGGGGAGACGCTTACTTCAAGCGTGGCCCTGGAAGAGACTTCCGACAGTACACCG	900
OY	1058	CCTTCAACATCCCGAGCATCAACAGAGAGACCCCGGAGCTCGCTACCAAGTACAAAGTGC	1117
Db	901	CCTTCAACATCCCGAGCATCAACAGAGAGACCCCGGAGCTCGCTACCAAGTACAAAGTGC	960
OY	1118	TGCCCCAGGGCTGAAAGGGCAAGCCCGACATTTTCAGAGCAAGTACCAAGATCTTGG	1177
Db	961	TGCCCCAGGGCTGAAAGGGCAAGCCCGACATTTTCAGAGCAAGTACCAAGATCTTGG	1020
OY	1178	AGCCCTTTCGCGCCCGCAACCCCGAGATTCGTGATTTTACAGAGCCCTCTGTACGTGGCA	1237
Db	1021	AGCCCTTTCGCAAGAGAAACCCCGACATTCGTGATTTTACAGAGCCCTCTGTACGTGGCA	1080
OY	1238	GCAGACTGAGAGTCCGACGACACCGCGCAAGATGAGAGAGCTGCGCAGACACTTGTGC	1297
Db	1081	GCAGACTGAGAGTCCGACGACACCGCAACAAATGAGAGAGCTGCGCAGACACTTGTGC	1140
OY	1298	GCTGGGGCTTTCACCAACCCCGCAAGAAAGCAACAAAGAGAGCCCTTCTGTCCCAT--	1355
Db	1141	GCTGGGGCTTTCACCAACCCCGCAAGAAAGCAACAAAGAGAGCCCTTCTGTGTGATGG	1200

QY	1356	-----GAGCTGCAACCCCGACAAGTGGACCGGTGACGCCATCTAGACTCAGCCGAGAAAGGA	1411
Db	1201	GCTACGAGCTGCACCCCGACAAGTGAACCGGTGACGCCCATCTATCTGCGCGAAGGAACA	1260
QY	1412	GCTGGAACCTGTGAACGACATCCAGAAAGCTGTGGCAAGCTGAATCTGGGCAAGCAGATCT	1471
Db	1261	GCTGGAACCTGTGAACGACATCCAGAAAGCTGTGGCAAGCTGAATCTGGGCAAGCAGATCT	1320
QY	1472	ACCCGCGATCAAGGTGGCGCAAGCTGTGCAAGCTGTGCGCGGCGCAAGGCTTGACCG	1531
Db	1321	ACGCGCGCATCAAGGTGAAGAGCTGTGCAAGCTGTGCGCGGCGCAAGGCTTGACCG	1380
QY	1532	ACATGTGGCCCTTACCGAAGAGGCCGAGCTGTGAGCTGGCTGGAATAACCGAGATCTCTGC	1591
Db	1381	AGGTGATCTCCCTTACCGAAGAGGCCGAGCTGTGAGCTGGCTGGAATAACCGAGATCTCTGA	1440
QY	1592	GCGAGCCCGTGTCAAGCGGTGTATCTATACGACCCCAACAGAGACCTGTGTGGCCGAGATCCAGA	1651
Db	1441	AGGAGCCCGTGTCAAGAGTGTATCTATACGACCCCAACAGAGACCTGTGTGGCCGAGATCCAGA	1500
QY	1652	AGCAGGGCCAGACGACAGTGGACCTTACCAAGATCTACAGAGACCTCTCAAGAACCTTGAA	1711
Db	1501	AGCAGGGCCAGAGGCCAGTGGACCTTACCAAGATCTACAGAGACCTCTCAAGAACCTTGAA	1560
QY	1712	CCGGCAGATACGCCAAGATGTGCAACCGCCACACCAACAGATGTGAAGACTTGACCGAG	1771
Db	1561	CCGGCAGATACGCCCGGCAATGCGCGGCGCCACACCAACAGATGTGAAGACTTGACCGAG	1620
QY	1772	CCGTGCAAGAAATGCGCATGTGAGAGCATCTGTATCTGTGGGCAAGACCCCAAGTTCCGCG	1831
Db	1621	CCGTGTCAAGAGTGAAGCACCGAAGCATCTGTATCTGTGGGCAAGATCTCCAGTTCCAGC	1680
QY	1832	TGCCCATTCAGAGAGAGACTCTGGGAGACTGTGTGAACGACTTACTGGCAGGCCACTTGA	1891
Db	1681	TGCCCATTCAGAGAGAGACTCTGGGAGCGCTGTGTGAATGTGAAGTACTGGCAGGCCACTTGA	1740
QY	1882	TCCCCGAGTGGGAATTTGTGTAAACCCCCCTCTGTGTGAAGCTGTGTATCCAGCTTGAGA	1951
Db	1741	TCCCCGAGTGGGAATTTGTGTAAACCCCCCTCTGTGTGAAGCTGTGTATCCAGCTTGAGA	1800
QY	1952	AGGAGCCCATCATCGGGCGCGGAGACTTTTACGTGTGAACGCGCGCGGCAACCGGAGAGCA	2011
Db	1801	AGGAGCCCATGTGTGGCGCGAGACCTTCTACGTGTGAACGCGCGCGGCAACCGGAGAGCA	1860
QY	2012	AGATCGGCAAGGCCGCTTACGTGACCGGACCGGGCGCGGCAAGAAATCGTGAAGCTTGACCG	2071
Db	1861	AGCTGGGCAAGGCCGCTTACGTGACCGGACCGGGCGCGGCAAGAAATCGTGAAGCTTGACCG	1920
QY	2072	AGACCAACCAACGAGAGACCGAGCTTGCAAGCCATCTCAGCTGGCCCTTGAGAGACGCGCA	2131
Db	1921	ACACCAACCAACGAGAGACCGAGCTTGCAAGCCATCTCAGCTGGCCCTTGAGAGACGCGCC	1980
QY	2132	GCGAGTGAATAATGTGTACCGGACAGGCCAGTACGCGCTGGGATATATCCAGGCCCAAGCCG	2191
Db	1981	TGGAGTGAATAATGTGTACCGGACAGGCCAGTACGCGCTGGGATATATCCAGGCCCAAGCCG	2040
QY	2192	ACAAAGCGGAGAGCGAGCTGTGTGAACCAAGATCATCTGAGCAGCTGATCAAGAAGAGAAAG	2251
Db	2041	ACAAAGCGGAGAGCGAGCTGTGTGAACCAAGATCATCTGAGCAGCTGATCAAGAAGAGAAAG	2100
QY	2252	TGTACTTGAAGTGGGTGCCCGCCACAAAGGCATCGGCGGCAACGAGCAGATTCAGAC	2311
Db	2101	TGTACTTGAAGCTGGGTGCCCGCCACAAAGGCATCGGCGGCAACGAGCAGATTCAGAC	2160
QY	2312	TGTGTGAGCAAGGGCATTCGCAAGTGTCTGTCTGTGAACGCGATCGATGGCGGCAATCTGTGA	2371
Db	2161	TGTGTGAGCGCGGGATTCGCAAGGTGTCTGTCTGTGAACGCGATCGATGGCGGCAATCTGTGA	2220
QY	2372	TCCTACAGTACATGGAAGCACTGTACGATGAGGCAAGCGCGGCGCCTTAGATTCGATTTAAAGC	2431
Db	2221	TCCTACAGTACATGGAAGCACTGTACGATGAGGCAAGCGCGGCGCCTTAGATTCGATTTAAAGC	2280
QY	2432	TTCCCGGGGCTAGCACCGGTGAATTC	2457

Db 2281 TTCCCGGGGCTAGACCGGTGAATTC 2306

RESULT 8
LOCUS CQ870574 2306 bp DNA linear PART 13-SRP-2004
DEFINITION Sequence 82 from Patent EP1433851.
ACCESSION CQ870574
VERSION CQ870574.1 GI:52000090
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS 1
Barnett, S., Zumege, J., Srivastava, I., Lian, Y., Hartog, K.,
Liu, H., Greer, C., Selby, M., and Walker, C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
1. .2306
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS (-). protod. Rptc. YM"

ORIGIN

Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
Best Local Similarity 93.2%; Pred. No. 3; 9e-165;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGGCCACGATGAGACTGCAACGAGCGCCAGCACTTCTTCGGG 229
Db 1 GCGGCGGCAAGAGCAACCAATGAAAGATTGCACTGAGAGCAAGGCTAATTTCTTCGGG 60

QY 230 AGGACTTGACCTTCCCGCAAGGCGCGAGTTCGCCAGCGAGCAAAACCGCGCA 289
Db 61 AGGACTTGACCTTCCCGCAAGGCGCGAGTTCGCCAGCGAGCAAAACCGCGCA 120

QY 290 ACAAGCCCAACAGCCGAGCTGCAAGTGGCGCGG-----ACAACCCCGAGAGAG 343
Db 121 ACAAGCCCAACAGCCGAGCTGCAAGTGGCGCGGCGAGCAACAGCTGAGAGAG 180

QY 344 CCGGCGCGAGGCGCGAGGCAACCTG-----AATTTCCCGAGATCAACCTGTGGAG 397
Db 181 CCGGCGCGAGGCGCGAGGCAACCTG-----AATTTCCCGAGATCAACCTGTGGAG 240

QY 398 GCGCCCTGAGATCAAGTGGCGCGAGATCAAGAGGCTGTGAGACACCGCG 457
Db 241 GCGCCCTGAGATCAAGTGGCGCGAGATCAAGAGGCTGTGAGACACCGCG 300

QY 458 CCGAGCAACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATATG 517
Db 301 CCGAGCAACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATATG 360

QY 518 GCGGCAATCGCGGCTTCAATGAAGTGGCGAGTGAACAGATCTGATGAAATCTGG 577
Db 361 GCGGCAATCGCGGCTTCAATGAAGTGGCGAGTGAACAGATCTGATGAAATCTGG 420

QY 578 GCAAGAGGCGATCGGCAACGCTGTGATCGGCGCCCAACCCCGTGAACATATGCGCG 637
Db 421 GCAAGAGGCGATCGGCAACGCTGTGATCGGCGCCCAACCCCGTGAACATATGCGCG 480

QY 638 ACAATGCTGACCAAGTGGCTGCAACCTTGAATTTCCCATCAAGCCCATGAGACGCTGC 697
Db 481 ACAATGCTGACCAAGTGGCTGCAACCTTGAATTTCCCATCAAGCCCATGAGACGCTGC 540

QY 698 CCGTGAAGCTGAAGCCCGGATGAGAGGCGCCCAAGTGAAGAGTGGCGCTGACGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGGATGAGAGGCGCCCAAGTGAAGAGTGGCGCTGACGAGG 600

QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGCGCAAGATCACCA 817
Db 601 AGAAGATCAAGGCGCTGAGATGTGACCGAATGAGAGAGGCGCAAGATCACCA 660

QY 818 AGATGCGCCCGAGAGACCCCTTACCAACCCCGTGTGCTCATCAAGAGAGACAGCA 877
Db 661 AGATGCGCCCGAGAGACCCCTTACCAACCCCGTGTGCTCATCAAGAGAGACAGCA 720

QY 878 CCAAGTGGCGCAAGTGGTGAATTCGCGAGCTGAACAAGCGCAACCGAGACTTCTGG 937
Db 721 CCAAGTGGCGCAAGTGGTGAATTCGCGAGCTGAACAAGCGCAACCGAGACTTCTGG 780

QY 938 AGTGCAGTGGGCAATCCCGCAACCCCGCGCTGAAGAGAGAGCGTGAACCTTTC 997
Db 781 AGTGCAGTGGGCAATCCCGCAACCCCGCGCTGAAGAGAGAGCGTGAACCTTTC 840

QY 998 TGAAGTGGGCGAGCGCTTCAAGCGTGGCCCTGAGACAGAGACTTCCGAGTACAGG 1057
Db 841 TGAAGTGGGCGAGCGCTTCAAGCGTGGCCCTGAGACAGAGACTTCCGAGTACAGG 900

QY 1058 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCAATCCGCTACAGTCAACGTCG 1117
Db 901 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCAATCCGCTACAGTCAACGTCG 960

QY 1118 TGCCTCAAGGCTGAGAGGCGAGCCCGAGATCTTCAAGGCGAGTGAACCAAGATCTGG 1177
Db 961 TGCCTCAAGGCTGAGAGGCGAGCCCGAGATCTTCAAGGCGAGTGAACCAAGATCTGG 1020

QY 1178 AGCCCTTCGCGCGCGCAACCCCGAGATGATGATCAACAGGCGCCCTGTAGTGGGA 1237
Db 1021 AGCCCTTCGCGCGCGCAACCCCGAGATGATGATCAACAGGCGCCCTGTAGTGGGA 1080

QY 1238 GCGACTGAGATCGGCGAGCAACCGCGCAAGTGAAGAGTGGCGAGCAACCTGCTGC 1297
Db 1081 GCGACTGAGATCGGCGAGCAACCGCGCAAGTGAAGAGTGGCGAGCAACCTGCTGC 1140

QY 1298 GCTGGGAGCTTCAACACCCCGCAAGAGACCAAGAGAGGCGCCCTTCTGCTCCAT-- 1355
Db 1141 GCTGGGAGCTTCAACACCCCGCAAGAGACCAAGAGAGGCGCCCTTCTGCTCCAT-- 1200

QY 1356 -----GAGCTGACCCCGCAAGTGAACGTCGAGCCCATGAGTGGCCGAGAGAG 1411
Db 1201 GCTGACAGTGAACCCCGCAAGTGAACGTCGAGCCCATGAGTGGCCGAGAGAG 1260

QY 1412 GCTGACCGTGAACGATCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCAAGCATCT 1471
Db 1261 GCTGACCGTGAACGATCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCAAGCATCT 1320

QY 1472 ACCCGGCAATCAAGTGGCGAGCTGTGCAAGCTGTGCGCGCGCAAGCCCTTGA 1531
Db 1321 ACCCGGCAATCAAGTGGCGAGCTGTGCAAGCTGTGCGCGCGCAAGCCCTTGA 1380

QY 1532 ACATGTGCGCTTGAACCGAGAGGCGAGCTGAAGTGGCCGAGAACCCCGAGATCTTC 1591
Db 1381 ACATGTGCGCTTGAACCGAGAGGCGAGCTGAAGTGGCCGAGAACCCCGAGATCTTC 1440

QY 1592 GCGAGCCGTGACAGGCGTGTACTACAGCCCGCAAGAGACTGTGGCCGAGATCAAG 1651
Db 1441 GCGAGCCGTGACAGGCGTGTACTACAGCCCGCAAGAGACTGTGGCCGAGATCAAG 1500

QY 1652 AGCAGGCGCAAGCAAGTGAACCTTACAGATCTTACAGAGGCCCTTCAAGAACCTGA 1711
Db 1501 AGCAGGCGCAAGCAAGTGAACCTTACAGATCTTACAGAGGCCCTTCAAGAACCTGA 1560

QY 1712 CCGGCAAGTGAACGCAAGTGGCGCAACCGCAACCAAGTGAAGAGCTGAACGAGG 1771
Db 1561 CCGGCAAGTGAACGCGAGTGGCGCGCGCAACCAAGTGAAGAGCTGAACGAGG 1620

QY 1772 CCGTGAAGATCGCATGAGAGCATGTGATCTTGGGCGAAGACCCCAAGTTCCGC 1831
Db 1621 CCGTGAAGATCGCATGAGAGCATGTGATCTTGGGCGAAGATCCCAAGTTCCGC 1680

542 TGCGCAGTACGACGAGATCTGTATCGATCTGCGCAGAAAGGCAATCGGCAACGGTGC 601
3500 TGCGCAGTACGACGAGATCTGTATCGATCTGCGCAGAAAGGCAATCGGCAACGGTGC 3559
602 TGAATGGGCCCCCCCCCGTGAATCATATGCGCGCAATGCTGATCCAGCTGGGCTGCA 661
3560 TGGTGGGCCCCCCCCCGTGAATCATATGCGCGCAATGCTGATCCAGATCGGCTGCA 3619
662 CCTGAATCTTCCCATCAAGCCCCCATCGAAGCGTGCCTGTAAGGCTGAAGCCGGGCAATG 721
3620 CCTGAATCTTCCCATCAAGCCCCCATCGAAGCGTGCCTGTAAGGCTGAAGCCGGGCAATG 3679
722 ACGGCCCCCAAGGTGAAGCAAGTGGCCCCCTGACCGAAGAAATCAAGGCTTGAACCGCA 781
3680 ACGGCCCCCAAGGTGAAGCAAGTGGCCCCCTGACCGAAGAAATCAAGGCTTGAAGGA 3739
782 TCTTGAAGAGATGGAAGAAAGGCGCAATCAACAAATCGGCCCCCGAAGAACCTCTTACA 841
3740 TCTTGAACCGAATGGAAGAAAGGCGCAATCAACAAATCGGCCCCCGAAGAACCTCTTACA 3799
842 ACAACCCCGGTGTTCGTCATCAAGAAAGAGACGACCAAGTGGCGCAAGCTGTGTGACT 901
3800 ACAACCCCGGTGTTCGTCATCAAGAAAGAGACGACCAAGTGGCGCAAGCTGTGTGACT 3859
902 TCCTGAGAGTGAACAAGCGCACCCAGAACTTCTGGAAGTGCAGCTGGGCAATCCCCCAAC 961
3860 TCCTGAGAGTGAACAAGCGCACCCAGAACTTCTGGAAGTGCAGCTGGGCAATCCCCCAAC 3919
962 CGGCGGCGCTTGAAGAAAGAAAGAGAGCTGACCTGTGTGAAGTGGGCGCAACGCTTACTTCA 1021
3920 CGGCGGCGCTTGAAGAAAGAAAGAGAGCTGACCTGTGTGAAGTGGGCGCAACGCTTACTTCA 3979
1022 GCGTCCCGCTGACGAGGACTTCCGCAAGTACACGCGCTTCAACATCCAGATCAACA 1081
3980 GCGTCCCGCTGACGAGGACTTCCGCAAGTACACGCGCTTCAACATCCAGATCAACA 4039
1082 ACGAGACCCCGCGCATCGCTACAGATCAACGTCGTGCGCCAGGCTGGAAGGCAAGC 1141
4040 ACGAGACCCCGCGCATCGCTACAGATCAACGTCGTGCGCCAGGCTGGAAGGCAAGC 4099
1142 CGAGCATCTTCAAGAGCAGATGACCAAGATCTGTGAAGCTTCCGCGCGCAACCCG 1201
4100 CGGCAATCTTCAAGAGCAGATGACCAAGATCTGTGAAGCTTCCGCGCAAGAACCCG 4159
1202 AGATGATGATCTACCA-----GACCCCGCTGATGATGAGGCAAGCTGAGATCGGCG 1255
4160 ACATGATGATCTACCAAGTACATGACACATCTGATGAGGCAAGCTGAGATCGGCG 4219
1256 AGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGAGGCTTCAACCAAC 1315
4220 AGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCTGAGGCTTCAACCAAC 4279
1316 CGGACAAAGAGACCAAGAAAGAGGCCCCCTTCTGCGCAT-----CGAGCTGACACCCG 1369
4280 CGGACAAAGAGACCAAGAAAGAGGCCCCCTTCTGAGATGAGGCTTCAAGAGCTGACACCCG 4339
1370 ACAAGTGAACGCTGACGACCATCGAGAGCTGCGCAAGAGAGAGCTGACCGTGAACGACA 1429
4340 ACAAGTGAACGCTGACGACCATCGAGAGCTGCGCAAGAGAGAGCTGACCGTGAACGACA 4399
1430 TCAGAGAGCTGAGGAGAGCTGAACTGAGGACGCGCAAGATCTTACCCCGCATCAAGGTGC 1489
4400 TCAGAGAGCTGAGGAGAGCTGAACTGAGGACGCGCAAGATCTTACCCCGCATCAAGGTGC 4459
1490 GCGAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCTTGAACGACATCTGTGCTTGAACCG 1549
4460 GCGAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCTTGAACGACATCTGTGCTTGAACCG 4519
1550 AGGAGGCGAGCTGAGAGCTGCGCGGCAAGACCGGAGATCTGTGCGCGCGCGCGCGCG 1609
4520 AGGAGGCGAGCTGAGAGCTGCGCGGCAAGACCGGAGATCTGTGCGCGCGCGCGCGCG 4579
1610 TGTACTACGACCCGACGACGAGATCTGTGCGCGAGATCCAGAGGAGGCGCACGACGAT 1669

4580 TGTACTACGACCCGACGAGAGGACCTGATCGCGCGAGATCCAGAGAGGAGCGGCGCAGT 4639
1670 GGAATCTACAGATCTTACAGAGAGGCTTCAAGAAACCTGTAAGACCGGCAAGTACCGCAAGA 1729
4640 GGAATCTACAGATCTTACAGAGAGGCTTCAAGAAACCTGTAAGACCGGCAAGTACCGCGCA 4699
1720 TGCGCACCGGCGCACCAAGACGATGAAGAGCTGACCGAGAGCGATGACAGAAATCGGCA 1789
4700 TGAAGGCGGCGCACCAAGACGATGAAGAGCTGACCGAGAGCGATGACAGAAATCGGCA 4759
1790 TGAAGAGATCTGTATCTGAGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGAGA 1849
4760 CCGAGAGATCTGTATCTGAGGCAAGACCCCAAGTTCAAGCTGCGCATCCAGAAAGAGA 4819
1850 CCGAGAGATCTGTATCTGAGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGAGA 1909
4820 CCGAGAGATCTGTATCTGAGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAAAGAGA 4879
1910 TGAACACCCCGCTGTGAGAGCTGTGATACAGCTGAGAGAGGCGCATCATCGGCG 1969
4880 TGAACACCCCGCTGTGAGAGCTGTGATACAGCTGAGAGAGGCGCATCATCGGCG 4939
1970 CCGAGACCTTCTACGATGAGAGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGGCT 2029
4940 CCGAGACCTTCTACGATGAGAGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGGCT 4999
2030 ACGTGAACCGAGCGGCGCGCGCAAGATCTGTGAGCTTGAACCGACCAACCAAGAGAGA 2089
5000 ACGTGAACCGAGCGGCGCGCGCAAGATCTGTGAGCTTGAACCGACCAACCAAGAGAGA 5059
2090 CCGAGCTGACAGGACATCCAGCTGCGCTGACAGAGACGCGGAGCGAGAGTGAACATCGTGA 2149
5060 CCGAGCTGACAGGACATCCAGCTGCGCTGACAGAGACGCGGAGCGAGAGTGAACATCGTGA 5119
2150 CCGAGAGCTGATGAGCGCTGTGAGCATGATCCAGGCTTCAACAGAGCGAGAGAGC 2209
5120 CCGAGAGCTGATGAGCGCTGTGAGCATGATCCAGGCTTCAACAGAGCGAGAGAGC 5179
2210 TGTGTAACCAATCATCGAGAGCTGATCAAGAAAGAAAGATGATCTGTGAGGCTGCG 2269
5180 TGTGTAACCAATCATCGAGAGCTGATCAAGAAAGAAAGATGATCTGTGAGGCTGCG 5239
2270 CCGGCGCAAGAGGAGCTGCGCGGCAAGAGAGATGACCAAGCTGTGAGCAAGGAGATCC 2329
5240 CCGGCGCAAGAGGAGCTGCGCGGCAAGAGAGATGACCAAGCTGTGAGCAAGGAGATCC 5299
2330 GCAAGTGTCTTCTGTGAGCGGATGATGAGCGGATGATGATCAACAGTA 2381
5300 GCAAGTGTCTTCTGTGAGCGGATGATGAGCGGATGATGATCAACAGTA 5351

RESULT 11
BD263706 2312 bp DNA linear PAT 17-JUL-2003
BD263706
LOCUS Improved expression of HIV polypeptides and production of
DEFINITION virus-like particles.
ACCESSION BD263706.1 GI:33073474
VERSION JP 2002533124-A/73.
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
JOURNAL Patent: JP 2002533124-A 73 08-OCT-2002;
COMMENT CHIRON CORP
OS Artificial Sequence
PN JP 2002533124-A/73
PD 08-OCT-2002

PF 30-DEC-1999 JP 2000591193
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471, PI
 SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIYASTAVA, YING LIAN, PI
 KARIN HARTOG,
 PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC
 C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
 C12N5/10,
 PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
 Description of Artificial Sequence: FS(-). protomod. Rtopc (+) FH
 Location/Qualifiers
 FT source 1..2312
 FT location/Qualifiers
 1..2312
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN

Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
 Best Local Similarity 92.9%; Pred. No. 2,4e-163;
 Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

170 GGCGAAGGAGGCGCACGATGAGGAGCTGACCGAGGCGGCGCACTTCTCCGG 229
 1 GCGGCGCGAAGAGCACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60
 230 AGGACCTGAGCTTCCCGGAGGCGGAGGCGGAGTTTCCGAGCGAGCAAGAACCGGCGCA 289
 61 AGGACCTGAGCTTCTCGAGAGGCGGAGGCGGAGTTTCAAGCGAGCGAGACCGGCGCA 120
 290 AAGGCGCGACGAGCGGAGGCTGAGGTGCGGCGG-----ACAACCCCGGAGGAGG 343
 121 AAGGCGCGACGAGCGGCGGAGGCTGAGGTGCGGCGGCGGAGCAACAAGCTGAGGAGG 180
 344 CGGCGGCGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGCGGAGCAACAAGCTGAGGAGG 397
 181 CGGCGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGCGGAGCAACAAGCTGAGGAGG 240
 398 GCGGCGGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGCGGAGCAACAAGCTGAGGAGG 457
 241 GCGGCGGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGCGGAGCAACAAGCTGAGGAGG 300
 458 CGGAGCGACCGGCTGAGGAGGAGTGAAGCTGCGGCGGAGGAGGCGGAGGCGGAGGAGTGA 517
 301 CGGAGCGACCGGCTGAGGAGGAGTGAAGCTGCGGCGGAGGAGGCGGAGGAGGAGTGA 360
 518 GCGGCGGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 577
 361 GCGGCGGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 420
 578 GCAAGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 637
 421 GCGGAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 480
 638 ACAATGTAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 697
 481 ACTGCTGTAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 540
 698 CCGTGAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 757
 541 CCGTGAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 600
 758 AGAAGATCAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 817
 601 AGAAGATCAAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 877
 818 AGAATGCGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 937
 661 AGAATGCGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 720
 878 CCAAGTGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 937

721 CCAAGTGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 780
 938 AGGTCAGCTGGGAGCTTCCCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 997
 781 AGGTCAGCTGGGAGCTTCCCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 840
 998 TGAAGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 1057
 841 TGAAGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 900
 1058 CTTTACCAATCCCGAGATCAACAAGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1117
 901 CTTTACCAATCCCGAGATCAACAAGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 960
 1118 TGGCCGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 1177
 961 TGGCCGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGCGGAGGAGTGA 1020
 1178 AGGCTTCGCGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1231
 1021 AGGCTTCGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1080
 1232 TGGGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1291
 1081 TGGGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1140
 1292 TGGGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1351
 1141 TGGGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1200
 1352 CGAT-----GAGGCGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1405
 1201 GAGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1260
 1406 AGGAGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1465
 1261 AGGAGAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1320
 1321 AGATCTACCGCGGAGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1380
 1526 TGAAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1585
 1381 TGAAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1440
 1586 TCGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1645
 1441 TCGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1500
 1646 TCGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1705
 1501 TCGTGGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1560
 1706 TGAAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1765
 1561 TGAAGGAGGAGGCGGAGGCGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1620
 1766 CCGAGGCGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1825
 1621 CCGAGGCGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1680
 1826 TCGGCTGCGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1885
 1681 TGAAGTGGCGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1740
 1886 CCGTGGAGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1945
 1741 CCGTGGAGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 1800
 1946 TGAAGGAGGAGGAGGAGGAGGAGGCTGAGGTGCGGCGGAGGAGGAGTGA 2005
 1801 TGAAGGAGGAGGAGGAGGAGGAGGCTTCTACGTGAGGAGGAGGAGGAGTGA 1860

QY	2006	GGACCAAGATGGGGAAGGCGCGCTACCTGACCCGACCGGGGGCGCGCAAGATCTGTAGCC	2065
Db	1861	AGACCCAAAGCTGAGCAAGGCGCGCTACCTGTACCCGACCGGGGCGCGCAAGAGTGTGAGCA	1920
QY	2066	TGACCGAGACCAACCAACGAAGAACCAGCTGCAGGCGCATCAGCTGAGCCCTGCAGAGCA	2135
Db	1921	TGCGCGAACACCAACCAACGAAGAACCAGCTGCAGGCGCATCAGCTGAGCCCTGCAGAGCA	1980
QY	2126	GGCGCAGCGAGGTGAACATCTGTACCCGACAGCCAGTATGCCCTTGGGCATCATCCAGGCC	2185
Db	1981	GGCGCCTGGAGGTGAACATCTGTACCCGACAGCCAGTATGCCCTTGGGCATCATCCAGGCC	2040
QY	2186	AGCCCGACAAAGCGAGAGCGAGCTGTGTAAACCAAGATCATCGAGCAGCTGATCAAGAAG	2245
Db	2041	AGCCCGACAAAGCGAGAGCGAGCTGTGTAAACCAAGATCATCGAGCAGCTGATCAAGAAG	2100
QY	2246	AGAAAGTTCACCTGAGCTGGGTGCCCGCCACAAGGAGCATCGCGCGGCAACGACGACGATCG	2305
Db	2101	AGAAAGTTCACCTGAGCTGGGTGCCCGCCACAAGGAGCATCGCGCGGCAACGACGAGGTGG	2160
QY	2306	ACAAAGCTGTGAGCAAGGCGCATTCGCGCAGAGTGCTGTTCTGTGACGGCATCGATGGCGCGCA	2365
Db	2161	ACAAAGCTGTGAGCGCCGCGCATTCGCGCAAGGTGCTGTTCTGTAAACGACATCGATGGCGCGCA	2220
QY	2366	TCTGTATTCACAGTACATGACGACCTGTACGTGTGGGACGCGCGGCGCCTTAGATTCGATT	2425
Db	2221	TCTGTATTCACAGTACATGACGACCTGTACGTGTGGGACGCGCGGCGCCTTAGATTCGATT	2280
QY	2426	AAAAGCTTCCGGGGCTAGCACCGGTGAATTC	2457
Db	2281	AAAAGCTTCCGGGGCTAGCACCGGTGAATTC	2312

RESULT 12					
LOCUS	CO870576	2312 bp	DNA	linear	PAT 13-SEP-2004
DEFINITION	Sequence 84 from Patent EPI433851.				
ACCESSION	CO870576				
VERSION	CO870576.1				
KEYWORDS	GI:5200092				
SOURCE	.				
ORGANISM	synthetic construct				
REFERENCE	synthetic construct				
AUTHORS	other sequences; artificial sequences.				
TITLE	1				
JOURNAL	Barnett, S., Zumegede, D., Srivastava, I., Lian, Y., Hartog, K., Liu, H., Greer, C., Selby, M. and Walker, C., Improved expression of HIV polypeptides and production of virus-like particles Patent: EP 1433851-A 84 30-JUN-2004; CHIRON CORPORATION (US)				
FEATURES	location/Qualifiers				
source	1..2312				
	/organism="synthetic construct"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:32630"				
	/note="Description of Artificial Sequence: FS(-).promod.RTopt(+)"				
ORIGIN					
Query Match	81.3%	Score 1998.4	DB 6	Length 2312	
Best Local Similarity	92.9%	Pred. No. 2.4e-163			
Matches 2147	Conservative 0	Missmatches 141	Indels 24	Gaps 4	
Db	170 GCGGCAAGAGGGCCACGACGATGAAGACTGACACCGCAGCGCCAACTCTTCGCGG	229			
	1 GCGGCGCGGAAGAGACCAAAATGAAGAATTGACTGAAGACAGGCTAATTTCTTCGCGG	60			
Db	230 AGGACTTGAGCTTCCCGCCAGGCGCAAGCGCCGAGTTCCCGAGCGAGACGACGCGCCA	289			
	61 AGGACTTGAGCTTCCCGCCAGGCGCAAGCGCCGAGTTCCCGAGCGAGACGACGCGCCA	120			
Db	290 AAGGCCCCACGACCGCGAGCTGCAAGTGTCGCGCGG-----ACAACCCCGCAGCGAG	343			

Db	121	ACAGCCCCACCCGCGGAGCTGCAGGTGTGTGGGCGGCGAAGAACACAGCCTTGAGCGAGG	180
Oy	344	CCGGCGCCGAGCGCCAGGAGCAACCTTG-----AACTTCCCCAGATCAACCTGTGGAGC	397
Db	181	CCGGCGCCGAGCCGCGAGGGGACCGGTGAGCTTCAACTTCCCCAGATCAACCTGTGGAGC	240
Oy	398	GCCCCCTGTGTGAGCATCAAGTGGGCGGCGCCAGATCAAGAGAGCCCTGTGTGACACCGGCG	457
Db	241	GCCCCCTGTGTGACCATCAGGATCGGCGGCGCAGCTCAAGAGAGCGCTGTCTGACACCGGCG	300
Oy	458	CCGACGACACCGTGTGTGAGGAGATGAGCGCTGCACCGGCGCAAGTGGAAACCCAGATGATTCG	517
Db	301	CCGACGACACCGTGTGTGAGGAGATGAACTTGCACCGGCGCAAGTGGAAACCCAGATGATTCG	360
Oy	518	GCGGACATCGGCGGCTTTCATCAAGTGTGCGCAGTACGACAGATCTGTATCGAGATCTTGC	577
Db	361	GCGGACATCGGCGGCTTTCATCAAGTGTGCGCAGTACGACAGATCTCGTGTGAGATCTTGC	420
Oy	578	GCAAGAAAGCCATCGGACAACGTGTGATCGGCCCCACCCCGTGAACATCATCGGCGCGCA	637
Db	421	GCCACAAGGCCATCGGACACGTGTGTGTGGGCCCCACCCCGTGAACATCATCGGCGCGCA	480
Oy	638	ACAGGCTGACCCGAGCTGTGGCTGACACCTTGAACCTTCCCATCAAGCCCATGAGAGCGGTGC	697
Db	481	ACGCTGTGACCCGACATCGGCTGTGCACCTTGAACTTCCCATCAAGCCCATGAGAGCGGTGC	540
Oy	698	CCGTGAAAGCTGAAAGCCCGGCAATGGAACGCGCCCAAGTGAAGAGAGCTGACCCCTGACCGAG	757
Db	541	CCGTGAAAGCTGAAAGCCCGGGAATGGAACGCGCCCAAGTGAAGAGAGCTGACCCCTGACCGAG	600
Oy	758	AGAAATCAAGAGCCCTTGACCGCCATCTTGCAAGAGATGAGAGAGAGGCGCAAGATCAACA	817
Db	601	AGAAATCAAGAGCCCTGTGTGTGAGATCTGACACCAAGATGAGAGAGAGGCGCAAGATCAACA	660
Oy	818	AGATCGGCGCCCGAGAAACCCCTTAACAACCCCGGTGTTCGATCAAAAGAGAGACGACA	877
Db	661	AGATCGGCGCCCGAGAAACCCCTTAACAACCCCGGTGTTCGATCAAAAGAGAGACGACA	720
Oy	878	CCAAGTGGGCGAAGCTGTGTGACTTTCGCGCAGCTGGAACAAGCGCACCGAGACTTCTGGG	937
Db	721	CCAAGTGGGCGAAGCTGTGTGACTTTCGCGCAGCTGGAACAAGCGCACCGAGACTTCTGGG	780
Oy	938	AGGTGAGCTGTGGGATCCCCACCCCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGGTGC	997
Db	781	AGGTGAGCTGTGGGATCCCCACCCCGCGCGCTGAAGAGAGAGAGAGCGGTGACCGGTGC	840
Oy	998	TGGAACGTGTGGGAGATGCTTACCTTACAGGTGCCTCTGACAGAGACTTTCGCAAGTACACCG	1057
Db	841	TGGAACGTGTGGGAGATGCTTACCTTACAGGTGCCTCTGACAGAGACTTTCGCAAGTACACCG	900
Oy	1058	CCTTACACATCCCGAGATCAACAAGAGACCCCGGCAATCCGCTACAGTACAAAGTGC	1117
Db	901	CCTTACACATCCCGAGATCAACAAGAGACCCCGGCAATCCGCTACAGTACAAAGTGC	960
Oy	1118	TGCCCAAGGCGTGAAGAGGCGAGCCCCCAGCATCTTTCAGAGAGAGATGACCAAGATCTTGG	1177
Db	961	TGCCCAAGGCGTGAAGAGGCGAGCCCCCAGCATCTTTCAGAGAGAGATGACCAAGATCTTGG	1020
Oy	1178	AGCCCTTTCGCGGCCGCAACCCGAGATCTGTATCTTACA-----GGCCCGCTGTATCG	1231
Db	1021	AGCCCTTTCGCGAGAGCAAAACCCCGACATGTGTATCTTACAGTACATGAGACGACTGTATCG	1080
Oy	1232	TGGGACAGGACCTGTGAATGGGCGCAGACCGCGCGCAAGATCGAGGAGCTGTGCGAGGACG	1291
Db	1081	TGGGACAGGACCTGTGAATGGGCGCAGACCGCGCGCAAGATCGAGGAGCTGTGCGAGGACG	1140
Oy	1292	TGCTGCGCTGTGGGCTTTCACCAACCCCGCAGAAAGAGCACAGAAAGAGACCCCTTCTGTGC	1351
Db	1141	TGCTGCGCTGTGGGCTTTCACCAACCCCGCAGAAAGAGCACAGAAAGAGACCCCTTCTGTGT	1200
Oy	1352	CCAT-----CGAGCTGACCCCGCAGAGTGAACGTGTGACAGCCCATCGAGCTTGCAGAGA	1405

Db 1201 GGATGGGCTACAGAGCTGACCCCGACAAATGGAACCGTGAAGCCATCATCTGCCCCGAGA 1260
QY 1406 AGGAGAGCTGGAACCGTGAACGATCCAGAAAGCTGGTGGCAAGCTGAATCTGGGCCAGCC 1465
Db 1261 AGGACAGCTGGAACCGTGAACGATCCAGAAAGCTGGTGGCAAGCTGAATCTGGGCCAGCC 1320
QY 1466 AGATTTACCCCGGATCAAGAGTGGCCGACAGCTGTGCAAGTGTGTGGCGGCCCAAGGCC 1525
Db 1321 AGATTTACCCCGGATCAAGAGTGGCCGACAGCTGTGCAAGTGTGTGGCGGCCCAAGGCC 1380
QY 1526 TGAACGACATGTGACCCCTGACCGAAGAGCGAGCTGAGCTGGCCGAGAACCGCGAGA 1585
Db 1381 TGAACGAGGTGATCCCTGACCGAAGAGCGAGCTGAGCTGGCCGAGAACCGCGAGA 1440
QY 1586 TCCTGCGGAGCCCGTGAACGAGCTGTACTACGACCCCGAAGAGCCTGTGGCCGAGA 1645
Db 1441 TCCTGAAGAGCCCGTGAACGAGCTGTACTACGACCCCGAAGAGCCTGTGGCCGAGA 1500
QY 1646 TCCAGAGAGGAGCCCAAGACAGTGAACCTTACCAATCTACCAAGAGCCCTTCAAGAAC 1705
Db 1501 TCCAGAGAGGAGCCCAAGAGCAGTGAACCTTACCAATCTACCAAGAGCCCTTCAAGAAC 1560
QY 1706 TGAAGACCGGAGATACGACCAAGATGCGACCGCCCAACCAAGAGCTGAAGAGCTGA 1765
Db 1561 TGAAGACCGGAGATACGACCGCCGATGCGGCGCCCAACCAAGAGCTGAAGAGCTGA 1620
QY 1766 CCGAGGCGCTGCAAGATGCGCATGAGAGCATCTGTGATCTGTGGGCAAGACCCCCAGT 1825
Db 1621 CCGAGGCGCTGCAAGATGCGCATGAGAGCATCTGTGATCTGTGGGCAAGATCCCAAGT 1680
QY 1826 TCCGCTGCGCATCCAGAGAGACCTTGGAGACCTGTGGAGCCAGCTACTGTGGCAGGCA 1885
Db 1681 TCAAGGTGCTTCCAGAGAGACCTTGGAGAGCCTGTGGAGCCTGTGGATGTGATCTGGCAGGCA 1740
QY 1886 CCGTGAATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGGTACAGC 1945
Db 1741 CCGTGAATCCCGAGTGGAGTTCGTGAACACCCCGCTGTGAAGCTGTGGTACAGC 1800
QY 1946 TGGAGAGAGGCCATCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCCGACACCGCG 2005
Db 1801 TGGAGAGAGGCCATCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCCGACACCGCG 1860
QY 2006 AAGACCAAGTCGAGAGGCGCGCTACGTGAACGACCGGCGCGCGAGAGAGTGTGAGCC 2065
Db 1861 AAGACCAAGTCGAGAGGCGCGCTACGTGAACGACCGGCGCGCGAGAGAGTGTGAGCA 1920
QY 2066 TGAACGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2125
Db 1921 TGAACGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 2126 GCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2185
Db 1981 GCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2186 AGCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2245
Db 2041 AGCCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2246 AGAAGGTATACCTGAGCTGGTGGCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2305
Db 2101 AGAAGGTATACCTGAGCTGGTGGCGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2306 ACAAGCTGTGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2365
Db 2161 ACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2366 TCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2425
Db 2221 TCGTATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2426 AAAAGCTCCCGGAGGCTAGACCGGTGAATTC 2457
Db 2281 AAAAGCTCCCGGAGGCTAGACCGGTGAATTC 2312

RESULT 13
AR373389 2312 bp DNA linear PAT 18-DEC-2003
LOCUS Sequence 84 from patent US 6602705.
DEFINITION AR373389
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source location/Qualifiers
1..2312
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
Best Local Similarity 92.9%; Pred. No.2.4e-163;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;
QY 170 GCGGCAAGAGGAGCCACAGATGAAGAGACTGACCGAGCGCCAGGACCACTTCTTCGCG 229
Db 1 GCGGCGGAGAGAGACCAATGAATGTGACTGAGAGAGAGGTATTTCTTCGCG 60
QY 230 AGAAGCTGAGCTTCCCGGAGGAGAGCGCGGAGTTCCTCCAGAGAGAGAGAGCGGCA 289
Db 61 AGAAGCTGAGCTTCCCGGAGGAGAGCGCGGAGTTCAGAGAGAGAGAGAGCGGCA 120
QY 290 AAGGCCCCACAGCCGCGAGCTGAGAGTGGCGGGG-----ACAACCCCGCAGGAG 343
Db 121 AAGGCCCCACAGCCGCGAGCTGAGAGTGGCGGGGCGAGAGACAGCTGAGGAG 180
QY 344 CCGGCGGAGGAGCCAGGAGCACTG-----ACTTCCCGAGATCACTGTGGAGC 397
Db 181 CCGGCGGAGGAGCCAGGAGCACTGAGCTTCACTTCCCGAGATCACTGTGGAGC 240
QY 398 GCCCCTGTGAGATCAAGAGTGGCGGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 457
Db 241 GCCCCTGTGAGATCAAGAGTGGCGGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY 458 CCGAGAGACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517
Db 301 CCGAGAGACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 518 GCGGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 577
Db 361 GCGGAGTGGGAGGCTTCAAGAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 578 GCAAGAGGCAATCGGACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 637
Db 421 GCAAGAGGCAATCGGACCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 638 AATGTGAGCCAGAGTGGGCTGACCTTGAATTTCCCATCAAGCCCATCGAGAGAGAGAG 697
Db 481 AATGTGAGCCAGAGTGGGCTGACCTTGAATTTCCCATCAAGCCCATCGAGAGAGAGAG 540
QY 698 CCGTGAAGCTGAAGCCCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
Db 541 CCGTGAAGCTGAAGCCCGGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
QY 758 AAGAGATCAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 817
Db 601 AAGAGATCAAGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 818 AGATGGGCCCCAGAGAACCCCTACACACCCCGTGTTCGCATCAAGAGAGAGAGAGAGAG 877


```

Db      ||| 661 AGATGAGCCCGGAGAAACCCCTACACACCCCGTGTGGCATCAAGAAAGAGGACAGCA 720
Qy      ||| 878 CCAATGGCGCAAGCTGGTGGATCTTCCGAGCTGAAACAAGGCAACCCAGATCTTGGG 937
Db      ||| 721 CCAATGGCGCAAGCTGGTGGATCTTCCGAGCTGAAACAAGGCAACCCAGATCTTGGG 780
Qy      ||| 938 AGATGAGCTGGGAGATCCCGCAACCCCGCGCTGAAAGAAAGAAAGAGCTGACCTGAC 997
Db      ||| 781 AGATGAGCTGGGAGATCCCGCAACCCCGCGCTGAAAGAAAGAAAGAGCTGACCTGAC 840
Qy      ||| 998 TGGAGTGGGCGACGCTTACCTTCAAGCTGACCTTGGACGAGGACTTCCGCAAGTACACG 1057
Db      ||| 841 TGGAGTGGGCGACGCTTACCTTCAAGCTGACCTTGGACGAGGACTTCCGCAAGTACACG 900
Qy      ||| 1058 CTTACCATCCCGAGCATCAACCAAGAACCCCGGACATCCGCTACCAAGTACCAAGTAC 1117
Db      ||| 901 CTTACCATCCCGAGCATCAACCAAGAACCCCGGACATCCGCTACCAAGTACCAAGTAC 960
Qy      ||| 1118 TGGCCCAAGGAGCTGAAAGGCAAGCCCGAGCATCTTCAAGAGCAGATGACCAAGATCTGG 1177
Db      ||| 961 TGGCCCAAGGAGCTGAAAGGCAAGCCCGAGCATCTTCAAGAGCAGATGACCAAGATCTGG 1020
Qy      ||| 1178 AGCCCTTCGCGGCGCCGCAACCCCGAGATCTGATCTTACCA-----GGCCCCCTGTAGC 1231
Db      ||| 1021 AGCCCTTCGCGGCAAGAGAACCCCGACATCTGATCTTACCAAGTACATGAGCAGCTGTAGC 1080
Qy      ||| 1232 TGGGCAAGGACCTTGAAGATTCGGGCAAGCAGCAGCCCAAGTCAAGAGAGCTTGGCAAGCAC 1291
Db      ||| 1081 TGGGCAAGGACCTTGAAGATTCGGGCAAGCAGCAGCCCAAGTCAAGAGAGCTTGGCAAGCAC 1140
Qy      ||| 1292 TGGCTGCTGGGAGCTTCAACCAACCCCGCAAGAGACCAAGAGAGGCCCCCTTCTGTAC 1351
Db      ||| 1141 TGGCTGCTGGGAGCTTCAACCAACCCCGCAAGAGACCAAGAGAGGCCCCCTTCTGTAC 1200
Qy      ||| 1352 CCAT-----CGAGTGCACCCCGCAAGTGGACCGTGGACGCTGACCTGACCTGACGAGA 1405
Db      ||| 1201 GGATGGGCTACGAGCTGCACCCCGCAAGTGGACCGTGGACGCTGACCTGACCTGACGAGA 1260
Qy      ||| 1406 AGGAGAGCTGCACCGTGAACCGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACAGC 1465
Db      ||| 1261 AGGAGAGCTGCACCGTGAACCGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACAGC 1320
Qy      ||| 1466 AGATTTACCCCGGAGCTCAAGAGTGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCCC 1525
Db      ||| 1321 AGATTTACCGCGGAGCTCAAGAGTGCAGCTGTGCAAGCTGTGCGCGGCGCAAGGCCC 1380
Qy      ||| 1526 TGAACGACATCGTGCCTCTGACCGAAGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGA 1585
Db      ||| 1381 TGAACGAGGTGATCCCTCTGACCGAAGAGGCGGAGCTGGAGCTGGCCGAGAACCCGCGAGA 1440
Qy      ||| 1586 TCCCTCGGAGGCGCGGAGCGAGCGGAGTACTAGAACCCCGAGAAAGAACTGTGGCGCGAGA 1645
Db      ||| 1441 TCCCTGAAGAGCGCGGAGCGAGCGGAGTACTAGAACCCCGAGAAAGAACTGTGGCGCGAGA 1500
Qy      ||| 1646 TCCAGAAAGAGGCGCGAGCGAGCTGACCTTACCAAGTCTTCAAGAGAGCCCTTCAAGAAC 1705
Db      ||| 1501 TCCAGAAAGAGGCGCGAGCGGAGCTGACCTTACCAAGTCTTCAAGAGAGCCCTTCAAGAAC 1560
Qy      ||| 1706 TGAAGACCGGCAAGTACCGCAAGATGCGGACCGGCCACCAACGACGTTGAAGAGAGCTGA 1765
Db      ||| 1561 TGAAGACCGGCAAGTACCGCGGAGTGGCGGCGGCCACCAACGACGTTGAAGAGAGCTGA 1620
Qy      ||| 1766 CCGAGGCGCTGCAAGAAATCGGCATGAGAGACATGTGATCTGGGGCAAGACCCCGCAAGT 1825
Db      ||| 1621 CCGAGGCGCTGCAAGAAATCGGCATGAGAGACATGTGATCTGGGGCAAGATCCCGCAAGT 1680
Qy      ||| 1826 TCCGCTGCGCATCCAGAAAGAGACCTGGAGACCTGGTGAACCGACCTATCTGAGAGGCCA 1885
Db      ||| 1681 TCCAGCTGCGCATCCAGAAAGAGACCTGGAGAGCTGGTGAATGAGATCTGAGAGGCCA 1740
Qy      ||| 1886 CTTGATCCCGAGTGGAGATTCTGTAACACCCCGCTGTGTGAAGCTGTGTGTACAGC 1945

```

```

Db      ||| 1741 CTTGATCCCGAGTGGAGATTCTGTGAACACCCCGCCCTGTGTGAAGCTGTGTACAGC 1800
Qy      ||| 1946 TGGAGAAAGAGCCCATCATTCGGGCGCGGAGACCTTCTAGTGAACGGCGCCGCAACCGCG 2005
Db      ||| 1801 TGGAGAAAGAGCCCATCATTCGTGGGCGCGGAGACCTTCTAGTGAACGGCGCCGCAACCGCG 1860
Qy      ||| 2006 AGACCAAGATCGGAGAGGCGCGGCTACGTGACCGACCGGAGGCGCGGAGAAAGATCTGTAGCC 2065
Db      ||| 1861 AGACCAAGCTGGGAGAGGCGCGGCTACGTGACCGACCGGAGGCGCGGAGAAAGTGTGTAGCA 1920
Qy      ||| 2066 TGAACGAGACCAACCAACCAAGAACCGAGCTGCAAGGCGCATTCAGCTGGCCCTGCAAGACA 2125
Db      ||| 1921 TGGCGGACACCAACCAAGAACCGAGCTGCAAGGCGCATTCAGCTGGCCCTGCAAGACA 1980
Qy      ||| 2126 GGGGCAAGAGTGAATCTGTGAACCGACAGCGAGTACCGCTGGGAGCATTCAGAGGCC 2185
Db      ||| 1981 GGGGCTGGAAGTGAATCTGTGAACCGACAGCGAGTACCGCTGGGAGCATTCAGAGGCC 2040
Qy      ||| 2186 AGCCGCAAGAGCGAGAGCGAGGAGCTGTGAACCAAGATCATGAGCAGCTGATCAAGAAAG 2245
Db      ||| 2041 AGCCGCAAGAGCGAGAGCGAGGAGCTGTGAACCAAGATCATGAGCAGCTGATCAAGAAAG 2100
Qy      ||| 2246 AGAAGTGTACCTGAGCTGGGTGCCCGCCCAAGAGGCGATGGCGGCAACGAGCAGATCG 2305
Db      ||| 2101 AGAAGTGTACCTGAGCTGGGTGCCCGCCCAAGAGGCGATGGCGGCAACGAGCAGATCG 2160
Qy      ||| 2306 ACAAGCTGTGAGAGCAAGGCGCATTCGCAAGAGTGTCTTCTGGAACGGCATTCGATGGCGGCA 2365
Db      ||| 2161 ACAAGCTGTGAGAGCGCGGCGCATTCGCAAGAGTGTCTTCTGGAACGGCATTCGATGGCGGCA 2220
Qy      ||| 2366 TCGTGTATTCACAGTACATGAGACGACCTGTACGTGGAGCGGCGGCTTGAAGATCATTT 2425
Db      ||| 2221 TCGTGTATTCACAGTACATGAGACGACCTGTACGTGGAGCGGCGGCTTGAAGATCATTT 2280
Qy      ||| 2426 AAAAGCTTCGCGGAGCTAGCACCGGTGAATTC 2457
Db      ||| 2281 AAAAGCTTCGCGGAGCTAGCACCGGTGAATTC 2312

RESULT 14
AX427936 9788 bp DNA linear PAT 20-JUN-2002
LOCUS AX427936
DEFINITION Sequence 174 from Patent WO0232943.
ACCESSION AX427936
VERSION AX427936.1 GI:21538023
KEYWORDS
SOURCE
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 174 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1. 9788
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
Query Match 81.1%; Score 1992.4; DB 6; Length 9788;
Best Local Similarity 91.2%; Pred. No. 4.8e-163;
Matches 2181; Conservative 0; Mismatches 181; Indels 30; Gaps 5;
Qy      ||| 14 TGGCGGAGGCGATGAGCGGAGCGACAGCGGCAATCTGATGAGGCGAGCACTTCA 73
Db      ||| 2961 TGGCGGAGGCGATGAGCGGAGCGACAGCGGCAATCTGATGAGGCGAGCACTTCA 3020
Qy      ||| 74 AGGCGCCCAAGCGCATCATCAAGTCTTCAACTGGCGGAGAGGAGGCGACATCGCGCGCA 133

```


2270 CGGCCCAAGGAGCATCGCGCGCAACGAGCATCGAAGCTGTGTGAGCAAGGCACTCC 2329
Db CGGCCCAAGGAGCATCGCGCGCAACGAGCATCGAAGCTGTGTGAGCAAGGCACTCC 5294
Qy 2330 GCAAGTCTGTTCTGTGACGCGCATCGATGCGCGCATGTGATCTTACCAAGTA 2381
Db 5295 GCAAGTCTGTTCTGTGACGCGCATCGAAGGCGCCAGAGGAGCAAGCA 5346

RESULT 15
AX427931 9169 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 169 from Patent WO232943.
DEFINITION AX427931
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS
SOURCE
ORGANISM
1. 9169
synthetic construct
synthetic construct
other sequences; artificial sequences.

REFERENCE
1 Huang, Y. and Nabel, G.J.
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 169 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1. 9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Plasmid pVR1012k/s containing HIV genes"

ORIGIN
Query Match 80.5%; Score 1978.4; DB 6; Length 9169;
Best Local Similarity 91.1%; Pred. No. 7,8e-162;
Matches 2178; Conservative 0; Mismatches 186; Indels 28; Gaps 6;

Qy 14 TGGCCGAGCCCATGAGCGGCGCACCGCAATCTGTATGAGCGGCAACTTCA 73
Db 2967 TGGCCGAGCCCATGAGCGGCGCACCGCAATCTGTATGAGCGGCAACTTCA 3026
Qy 74 AGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCGCCCATCGCCGCA 133
Db 3027 AGGCGCAAGCGCAT---CAAGTCTTCACTGCGGCAAGAGGCGCCCATCGCCGCA 3083
Qy 134 ACTGCGCGCGCGCGCGCAAGAGGCTGTGAAATGTGCGGCAAGAGGCGCCCATCATGA 193
Db 3084 ACTGCGCGCGCGCGCGCAAGAGGCTGTGAAATGTGCGGCAAGAGGCGCCCATCATGA 3143
Qy 194 AGGACTGACCGGAGCGCGCAAGGCACTTCTTCCGCGAGGACTGTGCTTCCCGAGGCA 253
Db 3144 AGGACTGACCGGAGCGCGCAAGGCTT-TTTTITAGGAAAGATGTGAGCTTCCCAAGGGA 3202
Qy 254 AGGCGCGGAGTCTCCCGAGCGAGAAACCGCGCAACAGCCCAACGAGCGCGAGTGC 313
Db 3203 AGGCGGAGGAAATTTCTTCAAGGAGCAAGAGGCGCAAGCGCCCAAGAGAGAGCTTC 3262
Qy 314 AGGTGCGGCGG-----CGACAAACCCCGCGAGGAGCGCGCGCGCGAGGCGCA--- 364
Db 3263 AGGTGCGGAGGAGCAACAACTCTCTTCAAGAGCGAGGCGCGCATGAGCAAGGAACTG 3322
Qy 365 ---CCCTGAATCTTCCCGAGATCACCTGTGCGAGCGCGCGCGTGTGAGCATCAAGTGC 421
Db 3323 TATCTTTAGCTTCCCTCGATGATCACTTTTGGAGAGAGACCTTGTGTACATTAAGATAG 3382
Qy 422 GCGGCGAGATCAAGAGGCGCTGTGACACCGGCGCGAGCAACACGCTGTGAGAGGA 481
Db 3383 GCGGCGAGATCAAGAGGCGCTTGAACACCGGCGCGAGCAACCGTGTGAGAGGA 3442
Qy 482 TGAAGCTGCGCGCAAGTGAAGGCGCAAGATGATGCGGCGCATCGCGGCTTCAATCAAG 541
Db 3443 TGAAGCTGCGCGCGCGTGAAGGCGCAAGATGATGCGGCGCATCGCGGCTTCAATCAAG 3502

Qy 542 TGGCCAGTACGACCAAGATCTGTATCGAGATCTGCGCGCAAGAGGCGCATCGGCAACGTC 601
Db 3503 TGGCCAGTACGACCAAGATCTGTATCGAGATCTGCGCGCAAGAGGCGCATCGGCAACGTC 3562
Qy 602 TGAATGCGCGCAACCCCGTGAACATCATCGGCGCGCAATCTGAACCAAGCTGAGTGA 661
Db 3563 TGAATGCGCGCAACCCCGTGAACATCATCGGCGCGCAATCTGAACCAAGCTGAGTGA 3622
Qy 662 CCTGAACCTTCCCATGAGCGCGCGCATCGAGACGTCGCGCGCGTGAAGCTGAAGCGCGGCA 721
Db 3623 CCTGAACCTTCCCATGAGCGCGCGCATCGAGACGTCGCGCGCGTGAAGCTGAAGCGCGGCA 3682
Qy 722 AGGCGCCCAAGTGAAGAGTGAAGCGCGCGCGCGCGCGAGATCAAGGCGCGTGAACGCGCA 781
Db 3683 AGGCGCCCAAGTGAAGAGTGAAGCGCGCGCGCGCGCGAGATCAAGGCGCGTGAACGCGCA 3742
Qy 782 TCTGCGAGAGATGAGAGAGGCGCAAGATGACCAAGATGAGCGCGCGAGAACCCCTTACA 841
Db 3743 TCTGCGAGAGATGAGAGAGGCGCAAGATGACCAAGATGAGCGCGCGAGAACCCCTTACA 3802
Qy 842 AACCCCGCGTGTGCGCATCAAGAGAGAGCAAGACCAAGTGTGCGCAAGCTGTGAGCT 901
Db 3803 AACCCCGCGTGTGCGCATCAAGAGAGAGCAAGACCAAGTGTGCGCAAGCTGTGAGCT 3862
Qy 902 TCGCGAGCTGAAGAGCGCACCGAGGACTTCTGCGAGTGAAGTGAAGTGAAGTGAAGTGA 961
Db 3863 TCGCGAGCTGAAGAGCGCACCGAGGACTTCTGCGAGTGAAGTGAAGTGAAGTGAAGTGA 3922
Qy 962 CGCGCGCGCTGAAGAGAGAGAGCGTGAACGTCGTCGAGAGTGAAGTGAAGTGAAGTGA 1021
Db 3923 CGCGCGCGCTGAAGAGAGAGAGCGTGAACGTCGTCGAGAGTGAAGTGAAGTGAAGTGA 3982
Qy 1022 GCGTGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 3983 GCGTGCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4042
Qy 1082 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 4043 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4102
Qy 1142 CGAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db 4103 CGAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4162
Qy 1202 AGATGTGATCTTACCA-----GGCGCGCGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1255
Db 4163 AGATGTGATCTTACCAAGTGAATGAACCAAGTGAATGAACCAAGTGAATGAACCAAGTGA 4222
Qy 1256 AGCACCGCGCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1315
Db 4223 AGCACCGCGCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4282
Qy 1316 CGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1369
Db 4283 CGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4342
Qy 1370 ACAAGTGAACGTCGAGCGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1429
Db 4343 ACAAGTGAACGTCGAGCGCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4402
Qy 1430 TCCAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1489
Db 4403 TCCAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4462
Qy 1490 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1549
Db 4463 GCGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4522
Qy 1550 AGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1609
Db 4523 AGAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4582
Qy 1610 TGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1669

```
Db 4583 TGTACTGAGACCCGACGAGAGGACTGATGCGGAGATCCAGAGCAAGGGCCAGGGCCACT 4642
QY 1670 GGAACCTACAGATCTACACGAGAGCCCTTCAGAACCTGAGAGCCGGCAAGTACGCCAAGA 1729
Db 4643 GGAACCTACAGATCTACACGAGAGCCCTTCAGAACCTGAGAGCCGGCAAGTACGCCCA 4702
QY 1730 TGGCGACCGGCCCAACCAACGAGCTGAGAGAGCTGACCGAGGCCCTGTCAGAAAGATCGCCA 1789
Db 4703 TGAAGGGGCGCCCAACCAACGAGCTGAGAGAGAGCTGACCGAGGCCCTGTCAGAAAGATCGCCA 4762
QY 1790 TGAAGAGCATGAGATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAAAGAGA 1849
Db 4763 CCGAGAGCATGAGATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAAAGAGA 4822
QY 1850 CCTGGAGAGACTGTGAGACCGACTACTGAGCAGGCGCACTGATCCCGAGTGGAGTTG 1909
Db 4823 CCTGGAGAGCCCTGTGTGAGACCGACTACTGAGCAGGCGCACTGATCCCGAGTGGAGTTG 4882
QY 1910 TGAACACCCCCCTGTGTGAAAGCTGTGATCCAGCTGAGAGAGAGCCCATATCGGCG 1969
Db 4883 TGAACACCCCCCTGTGTGAAAGCTGTGATCCAGCTGAGAGAGAGCCCATATCGGCG 4942
QY 1970 CCGAAGCCTTTACGTGAGACGGCGCCGCAACCGGAGACCAAGATCGGCAAGCGGCT 2029
Db 4943 CCGAAGCCTTTACGTGAGACGGCGCCGCAACCGGAGACCAAGCTGGGCAAGCGGCT 5002
QY 2030 ACGTGACGACCGGGGCGGAGAGATGAGCTGACCGAGACCAACCAACAGAGA 2089
Db 5003 ACGTGACGACCGGGGCGGAGAGTGTGCTGACCGAGACCAACCAACAGAGA 5062
QY 2090 CCGAGCTGAGGCGCATCGACTGAGCTTGCAGAGACGCGGCAAGAGGTGAACATCTGA 2149
Db 5063 CCGAGCTGAGGCGCATCGACTGAGCTTGCAGAGACGCGGCGTGAAGGTGAACATCTGA 5122
QY 2150 CCGACAGCCAGTACCGGCTGTGGGATGATCAAGGCCCAAGCCCAAGAGCGAGCGAGC 2209
Db 5123 CCGACAGCCAGTACCGGCTGTGGGATGATCAAGGCCCAAGAGCGAGCGAGC 5182
QY 2210 TGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGC 2269
Db 5183 TGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGC 5242
QY 2270 CCGCCCAAGAGGCGATCGGCGGCAACGAGCAGATCGAACAAGCTGTGAGCAAGGCGATCC 2329
Db 5243 CCGCCCAAGAGGCGATCGGCGGCAACGAGCAGATCGGCGCTGTGAGCGCGGCGATCC 5302
QY 2330 GCAAGGTGCTTCTGTGAGCGGATGATGAGCGGCGATCGTGAATCTACAGTA 2381
Db 5303 GCAAGGTGCTTCTGTGAGCGGATGATGAGCGGCGCGAGGCGGAGGAGCAAGAGA 5354
```

Search completed: December 30, 2005, 19:16:36
Job time : 11936.2 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1293.51 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457
Sequence: 1 gtcgacgcaccacatgcccga.....gggtacgaccggtgaattc 2457

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*
- 14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	ABL39961 Synthetic
2	2457	100.0	2457	12	ADM7376 HIV-1 pol
3	2443.4	99.4	2445	8	ACA03546 Synthetic
4	2443.4	99.4	2445	10	ADCl3264 DNA of HI
5	2436.4	99.2	2453	6	ABL39960 Synthetic
6	2436.2	99.2	2453	12	ADM73765 HIV-1 pol
7	2436.2	99.1	3930	10	ADCl3231 DNA of HI
8	2434.8	99.1	3930	10	ADCl3232 DNA of HI
9	2434.8	99.1	5184	8	ACA03591 Synthetic
10	2434.8	99.1	5184	8	ADCl3279 DNA of HI
11	2428.6	98.8	2457	8	ACA03547 Synthetic
12	2428.6	98.8	2457	10	ADCl3265 DNA of HI
13	2415.4	98.3	2469	6	ABL39959 Synthetic
14	2415.4	98.3	2469	12	ADM73764 HIV-1 pol
15	2404.4	97.9	3531	10	ADCl3234 DNA of HI
16	2404.4	97.8	2457	8	ACA03548 Synthetic
17	2403.4	97.8	2457	10	ADCl3266 DNA of HI
18	2403.4	97.8	3537	10	ADCl3236 DNA of HI
19	2402.8	97.8	3537	10	ADCl3236 DNA of HI

20	2401.8	97.8	5145	8	ACA03521 Synthetic
21	2401.8	97.8	5145	10	ADCl3233 DNA of HI
22	2391.8	97.3	3538	10	ADCl3235 DNA of HI
23	2343.4	95.4	3607	8	ACA03551 Synthetic
24	2343.4	95.4	3607	10	ADCl3269 DNA of HI
25	2325.2	94.6	3597	8	ACA03549 Synthetic
26	2325.2	94.6	3597	10	ADCl3267 DNA of HI
27	2295.8	93.4	3624	8	ACA03550 Synthetic
28	2295.8	93.4	3624	10	ADCl3268 DNA of HI
29	2136	86.9	2460	8	ACA03541 Synthetic
30	2136	86.9	2460	8	ACC78505 HIV p2Pol
31	2135	86.9	3564	8	ACC78488 HIV GagPo
32	2135	86.9	3564	8	ACC78489 HIV GagPo
33	2134.4	86.9	4716	8	ACA03522 Synthetic
34	2134.4	86.9	4716	10	ADCl3238 DNA of HI
35	2130.6	86.7	3999	8	ACC78484 HIV GagCo
36	2129	86.7	3999	8	ACC78485 HIV GagCo
37	2129	86.7	3999	8	ACC78486 HIV GagCo
38	2129	86.7	5283	8	ACA03584 Synthetic
39	2129	86.7	5283	8	ACC78529 HIV TatRe
40	2128.8	86.6	4713	8	ACA03592 Synthetic
41	2128.8	86.6	4713	10	ADCl3280 DNA of HI
42	2128.4	86.6	3462	10	ADCl3237 DNA of HI
43	2115.2	86.1	2466	8	ACA03542 Synthetic
44	2115.2	86.1	2466	8	ACC78506 HIV p2Pol
45	2103	85.6	3735	8	ACA03545 Synthetic

ALIGNMENTS

RESULT 1	ABL39961	standard; DNA; 2457 BP.
ID	ABL39961	standard; DNA; 2457 BP.
XX	ABL39961;	
AC	ABL39961;	
XX	15-MAY-2002 (first entry)	
DT	Synthetic construct PR97SYMM SEQ ID NO:32.	
XX	Human immunodeficiency virus type C; antigenic HIV type C protein;	
KW	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;	
XX	immunostimulant; gene therapy; gene; ds.	
OS	Human immunodeficiency virus; type C.	
XX	Synthetic.	
XX	WO200204493-A2.	
XX	17-JAN-2002.	
PD	05-JUL-2001; 2001WO-US021241.	
XX	05-JUL-2001; 2001US-00610313.	
PR	(CHIR) CHIRON CORP.	
XX	(UYST-) UNIV STELLENBOSCH.	
PA	Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	
XX	WPI; 2002-154920/20.	
DR	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	
XX	applications including DNA immunization or generation of packaging cell	
PT	lines, particularly in gene therapy.	
PT	Claim 1; Fig 10; 233pp; English.	
XX	The present invention describes expression cassettes comprising a	
PS	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	
XX	type C polypeptides. The expression cassettes comprise any of the HIV	
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	

(I). *I* have immunostimulant activity and can be used in gene therapy. The HIV type C polynucleotides are useful in applications including DNA immunisation, generation of packaging cell lines, and production of HIV type C proteins. The polynucleotides are particularly useful in gene therapy and DNA immunisation applications. A515962 to A5L40054 and A5B06204 to A5B06215 represent sequences used in the exemplification of the present invention

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match	Score	DB	Length
100.0%	2457	6	2457

Matches 2457; Conservative 0; Mismatches 0;

1 GTGACGCCACCATGGCCGAGGCCATGAGCCAGGCCACCAGGCCAACATCTGATGCAG 60

Db	1	GTGAGGCCACCAATGGCCGAGGCGATGAGCCAGGCCACAGGCCCAACATCTGATGCGAG	60
OY	61	CGACGAACCTTCAAGGGGCCCCAAAGCGCATCATCAATGTCTTCACTTGGCGCAAGAGGGC	120
Db	61	CGACGAACCTTCAAGGGGCCCCAAAGCGCATCATCAATGTCTTCACTTGGCGCAAGAGGGC	120
OY	121	CACATTCGCCCGCACTGCGCGCGCCCCCGCGCAAGAAAGGGCTGCGAAAGTGCGGCAAGAG	180
Db	121	CACATTCGCCCGCACTGCGCGCGCCCCCGCGCAAGAAAGGGCTGCGAAAGTGCGGCAAGAG	180
OY	181	GGCCCAACAGATAGGAAGGACTGCACCGAGCGCCAGAGGCCCAATCTTCTTCGCGAGGACCTTGCC	240
Db	181	GGCCCAACAGATAGGAAGGACTGCACCGAGCGCGAGAGGCCCAATCTTCTTCGCGAGGACCTTGCC	240
OY	241	TTCCCGCAGGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAAGACCGCGCAACAGCCCCCACC	300
Db	241	TTCCCGCAGGGGCAAGGCCCGCGAGTTCCCGAGCGAGCAAGACCGCGCAACAGCCCCCACC	300
OY	301	AGCCGCGAGCTCAAGGTGCGCGCGGCAACACCCCGGAGGAGGGCGCGCGCGAGCGCCAG	360
Db	301	AGCCGCGAGCTCAAGGTGCGCGCGGCAACACCCCGGAGGAGGGCGCGCGCGAGCGCCAG	360
OY	361	GGCACCTTGAACTTCCCGCAGATCAACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG	420
Db	361	GGCACCTTGAACTTCCCGCAGATCAACCTGTGGCAGCGCCCCCTGTGTAGCATCAAGGTG	420
OY	421	GGCGGCGCAGATCAAGAGGGGCCCTGTGTGACACCGCGCGCGAGCAACCGTGTGTGAGAG	480
Db	421	GGCGGCGCAGATCAAGAGGGGCCCTGTGTGACACCGCGCGCGAGCAACCGTGTGTGAGAG	480
OY	481	ATGAGCCTTGCCTCGGCAGATGGAAGGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAG	540
Db	481	ATGAGCCTTGCCTCGGCAGATGGAAGGCCCAAGATGATCGCGCGCATCGCGGCTTCATCAAG	540
OY	541	GTCGCGCAGTACGACCAAGTCTGTATGAGATTTGCGCGCAAGAGCCATCGGCACCGTG	600
Db	541	GTCGCGCAGTACGACCAAGTCTGTATGAGATTTGCGCGCAAGAGCCATCGGCACCGTG	600
OY	601	CTGATTCGAGCCCAACCCCGCGTGAACATCATTCGCGCGCAACATGCTGAACCCAGCTGGGCTGC	660
Db	601	CTGATTCGAGCCCAACCCCGCGTGAACATCATTCGCGCGCAACATGCTGAACCCAGCTGGGCTGC	660
OY	661	ACCTTGAACTTCCCATCAAGCCCTCATGAGCCGTGCGCGGTGAAGCTGAAGCCCGCGCATG	720
Db	661	ACCTTGAACTTCCCATCAAGCCCTCATGAGCCGTGCGCGGTGAAGCTGAAGCCCGCGCATG	720
OY	721	GACGGGCCCAAGGTGAAGATGCGCCCGTGAACCGAAGAAAGATTCAAAGGCCCTGAACCGCC	780
Db	721	GACGGGCCCAAGGTGAAGATGCGCCCGTGAACCGAAGAAAGATTCAAAGGCCCTGAACCGCC	780
OY	781	ATCTCGAGAGATGAGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTAC	840
Db	781	ATCTCGAGAGATGAGAGAGGAGGCGCAAGATCAACCAAGATCGGCCCGCGAGAACCCCTAC	840
OY	841	AACACCCCGGTGTTGGCATCAAGAAAGAGACGACCAAGTGGCGCAAGCTGTGTGAC	900
Db	841	AACACCCCGGTGTTGGCATCAAGAAAGAGACGACCAAGTGGCGCAAGCTGTGTGAC	900

QY	901	TTCCGCGAGCTGAAACAAGCGCAACCCAGACTTTCTGGAGAGTGGAGCTGGAGATCCCCAC	960
Db	901	TTCCGCGAGCTGAAACAAGCGCAACCCAGACTTTCTGGAGAGTGGAGCTGGAGATCCCCAC	960
QY	961	CCCGCGAGCTGAAAGAAAGAAAGAGCGTGAACCGTGTGAGAGCTGGAGAGAGCTACTTTC	1020
Db	961	CCCGCGAGCTGAAAGAAAGAAAGAGCGTGAACCGTGTGAGAGCTGGAGAGAGCTACTTTC	1020
QY	1021	AGCGTGCCTCTGAGAGAGACTTTCGCGAAGTACACCGCTTACATTCGCCAGATCAAC	1080
Db	1021	AGCGTGCCTCTGAGAGAGACTTTCGCGAAGTACACCGCTTACATTCGCCAGATCAAC	1080
QY	1081	AACGAGACCCCGGACATCCGCTACAGTACAACTGTCTGACCCGAGGCTGGAAAGGAGAC	1140
Db	1081	AACGAGACCCCGGACATCCGCTACAGTACAACTGTCTGACCCGAGGCTGGAAAGGAGAC	1140
QY	1141	CCGAGCATCTTTCAGAGCAGATACCAAGATCTTGAGACCTTTCGCGCTCGCAACCC	1200
Db	1141	CCGAGCATCTTTCAGAGCAGATACCAAGATCTTGAGACCTTTCGCGCTCGCAACCC	1200
QY	1201	GAGATCGATCTACAGAGCCCTCTGTACGTGGGAGCGACCTGGAGATCGAGCAGAC	1260
Db	1201	GAGATCGATCTACAGAGCCCTCTGTACGTGGGAGCGACCTGGAGATCGAGCAGAC	1260
QY	1261	CGCGCCAGATCGAGAGCTGCGAGCAGCTGCGCTGGGAGCTTCAACCAACCCGAC	1320
Db	1261	CGCGCCAGATCGAGAGCTGCGAGCAGCTGCGCTGGGAGCTTCAACCAACCCGAC	1320
QY	1321	AAGAGAGCAGAGAGAGAGCCCTCTCTGTGCCATCGAGCTGGACCCCGACAATGGAC	1380
Db	1321	AAGAGAGCAGAGAGAGAGCCCTCTCTGTGCCATCGAGCTGGACCCCGACAATGGAC	1380
QY	1381	GTGCAAGCCATCGAGCTCCCGAGAGAGAGAGCTGGACCTGTGAACGATCAAGAACTGT	1440
Db	1381	GTGCAAGCCATCGAGCTCCCGAGAGAGAGAGCTGGACCTGTGAACGATCAAGAACTGT	1440
QY	1441	GTGGGCAAGCTGAACCTGGGCAAGCTTACCCCGGACTCAAGTGGTGGCGCAAGCTGTGC	1500
Db	1441	GTGGGCAAGCTGAACCTGGGCAAGCTTACCCCGGACTCAAGTGGTGGCGCAAGCTGTGC	1500
QY	1501	AAGCTGTGGGCGGCGCGCAAGGCTCTGACGACATCGTGGCCCTGACCGAGAGAGGCGAG	1560
Db	1501	AAGCTGTGGGCGGCGCGCAAGGCTCTGACGACATCGTGGCCCTGACCGAGAGAGGCGAG	1560
QY	1561	CTGAGAGCTGGCGAGAACCGGAGATCTCTGGCGAGCCCGTGCACCGCGTGTACTTACGAC	1620
Db	1561	CTGAGAGCTGGCGAGAACCGGAGATCTCTGGCGAGCCCGTGCACCGCGTGTACTTACGAC	1620
QY	1621	CCGAGCAAGAGACTGTGTGGCGAGATCTCAAGAGCAGAGGCGCAGACCAATGGACCTTACAG	1680
Db	1621	CCGAGCAAGAGACTGTGTGGCGAGATCTCAAGAGCAGAGGCGCAGACCAATGGACCTTACAG	1680
QY	1681	ATTCTACAGAGAGCCCTTCAAGACCTTGAAACCGGCAAGTACGCGCAAGATGCGCACCGAC	1740
Db	1681	ATTCTACAGAGAGCCCTTCAAGACCTTGAAACCGGCAAGTACGCGCAAGATGCGCACCGAC	1740
QY	1741	CACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGCAAGAAATCGCATGAGAGCCTAC	1800
Db	1741	CACACCAACGAGTGAAGCAGCTGACCGAGGCGGTGCAAGAAATCGCATGAGAGCCTAC	1800
QY	1801	GTGATCTGGGCGAAGACCCCGCAAGTTCGCGCTGGCCATCCAGAAAGAGACCTGGGAGAGCC	1860
Db	1801	GTGATCTGGGCGAAGACCCCGCAAGTTCGCGCTGGCCATCCAGAAAGAGACCTGGGAGAGCC	1860
QY	1861	TGGTGAACGACTACTGGCAGGCGACCTGTGATCCCGAGTGGAGATTCTGTAAACACCCC	1920
Db	1861	TGGTGAACGACTACTGGCAGGCGACCTGTGATCCCGAGTGGAGATTCTGTAAACACCCC	1920
QY	1921	CCCTGTGTGAACCTGTGTATCCAGCTTGAAGAGAGCCCATCATCTGGGCGCGAAGCTTTC	1980
Db	1921	CCCTGTGTGAACCTGTGTATCCAGCTTGAAGAGAGCCCATCATCTGGGCGCGAAGCTTTC	1980

polypeptides, useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, especially humans.

Example 1; Fig 34; 262pp; English.

The invention describes a synthetic polynucleotide encoding 2 or more immunogenic HIV polypeptides, where at least 2 of the polypeptides are derived from different HIV subtypes. The polynucleotide is useful for immunization, generation of packaging cell lines, or production of HIV polypeptides. The polynucleotide and its encoded proteins are useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, or for inducing neutralising antibodies against HIV. The gene delivery vector comprising the polynucleotide is also useful for DNA immunisation of, or for generating an immune response (e.g. a humoral or cellular immune response) in, a subject such as a mammal, particularly a human. This sequence encodes a human immunodeficiency virus immunogenic peptide

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 8; Length 2445;

Best Local Similarity 100.0%; Pred. No. 5.3e-294; Mismatches 1; Indels 0; Gaps 0;

Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 7 GCCACCATGCGCGAGCCATGAGCCAGCGCCACATCTCTGATGACGGCAGC 66
DB 1 GCCACCATGCGCGAGCCATGAGCCAGCGCCACATCTCTGATGACGGCAGC 60
QY 67 AACTTCAGAGGCGCCCAAGGCGATCATCAAGTCTTCAATGCGGCAAGAGGCGCCATC 126
DB 61 AACTTCAGAGGCGCCCAAGGCGATCATCAAGTCTTCAATGCGGCAAGAGGCGCCATC 120
QY 127 GCCCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 186
DB 121 GCCCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 187 CAGATGAAGAGATGCAACGAGCGCCAGCCACTTCTTCGCGAGGACCTTGGCTTCC 246
DB 181 CAGATGAAGAGATGCAACGAGCGCCAGCCACTTCTTCGCGAGGACCTTGGCTTCC 240
QY 247 CAGGCGAGAGCGCGCGAGTTCCTCCAGCGAGCGAGCGCGCGCGCGCGCGCGCGCG 306
DB 241 CAGGCGAGAGCGCGCGAGTTCCTCCAGCGAGCGAGCGCGCGCGCGCGCGCGCGCG 300
QY 307 GAGCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
DB 301 GAGCTGCAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 367 CTGAACCTTCCCGCAGATCACTCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGGCG 426
DB 361 CTGAACCTTCCCGCAGATCACTCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGGCG 420
QY 427 CAGATCAAGAGAGCGCTGTGCGACCGCGCGCGCGCGCGCGCGCTGTGAGGAGATGAC 486
DB 421 CAGATCAAGAGAGCGCTGTGCGACCGCGCGCGCGCGCGCGCGCTGTGAGGAGATGAC 480
QY 487 CTGCGCGGAGAGTGAAGCGCGAGAGATGCGCGGAGATGCGCGGCTTCAAGAGTGGCG 546
DB 481 CTGCGCGGAGAGTGAAGCGCGAGAGATGCGCGGAGATGCGCGGCTTCAAGAGTGGCG 540
QY 547 CAGTACGACGAGATCTGTATCGAGATCTGTGCGGAGAGAGCGCATCGGCACTGGCTGATC 606
DB 541 CAGTACGACGAGATCTGTATCGAGATCTGTGCGGAGAGAGCGCATCGGCACTGGCTGATC 600
QY 607 GCGCCCGACCGCGGTGAACATCATCGGCGCGCGCGCGCGCGCGCGCTGTGAGCGCTTG 666
DB 601 GCGCCCGACCGCGGTGAACATCATCGGCGCGCGCGCGCGCGCGCGCTGTGAGCGCTTG 660
QY 667 AACTTCCCGCATGCGCGCGCATCGAGACCGTGGCGGTGAAGTGAAGCGCGCGCGCATGAGCG 726
DB 661 AACTTCCCGCATGCGCGCGCATCGAGACCGTGGCGGTGAAGTGAAGCGCGCGCATGAGCG 720
```

```

QY 727 CCAAGTGAAGCAGTGGCGCGCTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTGC 786
DB 721 CCAAGTGAAGCAGTGGCGCGCTGACCGAGGAGAGATCAAGGCGCTGACCGCATCTGC 780
QY 787 GAGGAGATGAGAGAGAGGCGAGATCAACGATGCGCGCGAGAACCCCTTCAACAAC 846
DB 781 GAGGAGATGAGAGAGAGGCGAGATCAACGATGCGCGCGAGAACCCCTTCAACAAC 840
QY 847 CCGGTGTTCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
DB 841 CCGGTGTTCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 907 GAGCTGAACAGCGCACCGAGACTTCTGGAGAGTGAAGTGGAGATCCCGACCGCGCG 966
DB 901 GAGCTGAACAGCGCACCGAGACTTCTGGAGAGTGAAGTGGAGATCCCGACCGCGCG 960
QY 967 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
DB 961 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1027 CCGCTGAGCAGAGACTTCTCGAGAGTACACCGCTTACCATCCAGCATCAACAGAG 1086
DB 1021 CCGCTGAGCAGAGACTTCTCGAGAGTACACCGCTTACCATCCAGCATCAACAGAG 1080
QY 1087 ACCCGCGCATCCGCTTACAGTGAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1146
DB 1081 ACCCGCGCATCCGCTTACAGTGAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1147 ATCTTCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCACCCCGAGATC 1206
DB 1141 ATCTTCAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCACCCCGAGATC 1200
QY 1207 GTGATCTACAGAGCGCGCGCTGTACGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
DB 1201 GTGATCTACAGAGCGCGCGCGCTGTACGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1267 AAGATCGAGAGAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
DB 1261 AAGATCGAGAGAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1327 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
DB 1321 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1387 CCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
DB 1381 CCCATCGAGCTGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1447 AAGCTGAACCTGGGCGAGAGATCTACCCCGGAGATCAAGGTGGCGCGCTGTGCAAGCTG 1506
DB 1441 AAGCTGAACCTGGGCGAGAGATCTACCCCGGAGATCAAGGTGGCGCGCTGTGCAAGCTG 1500
QY 1507 CTGCGCGGCGCGAGAGCGCGCTGACCGAGATCGAGCGCGCTGACCGAGAGAGCGGAG 1566
DB 1501 CTGCGCGGCGCGAGAGCGCGCTGACCGAGATCGAGCGCGCTGACCGAGAGAGCGGAG 1560
QY 1567 CTGCGCGAGAGAGCGCGAGATCTGTGCGAGAGCGCGCTGACCGAGAGAGCGGAG 1626
DB 1561 CTGCGCGAGAGAGCGCGAGATCTGTGCGAGAGCGCGCTGACCGAGAGAGCGGAG 1620
QY 1627 AAGAGCTGTGGCGCGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
DB 1621 AAGAGCTGTGGCGCGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1687 CAGAGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
DB 1681 CAGAGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1747 AAGCAGTGAAGCAGAGTGAACGAGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806
DB 1741 AAGCAGTGAAGCAGAGTGAACGAGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1807 TGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
```

DB 1801 TGGGGCAGAGCCCGCAAGTTCGCGCTGCCCATCCGAGAGAGAGACTGGGAGACCTGGTGG 1860
QY 1867 ACCGACTCTGCGAGGCGCAACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTG 1926
DB 1861 ACCGACTCTGCGAGGCGCAACCTGGATCCCGAGTGGGAGTTCTGTAACACCCCGCCCTG 1920
QY 1927 GTGAAGCTGTGGTACCACTGAGAGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACCTG 1986
DB 1921 GTGAAGCTGTGGTACCACTGAGAGAGAGAGCCCATCATCGCGCGCGAGACCTTCTACCTG 1980
QY 1987 GACGCGCGCGCGCAACCGCGAGACCAAGATCCGCAAGCGCGCTGACGTGACCGAGCGGCGC 2046
DB 1981 GACGCGCGCGCGCAACCGCGAGACCAAGATCCGCAAGCGCGCTGACGTGACCGAGCGGCGC 2040
QY 2047 CGGCGAAGATCTGTAGCTGACCGAGACCAAGATCCGCAAGAGACCGAGCTGACGCGCTC 2106
DB 2041 CGGCGAAGATCTGTAGCTGACCGAGACCAAGATCCGCAAGAGACCGAGCTGACGCGCTC 2100
QY 2107 CAGCTGGCGCTGCGAGACGAGCGCGAGGTTGAACATCTGTGACCGAGCGCGAGTACGCC 2166
DB 2101 CAGCTGGCGCTGCGAGACGAGCGCGAGGTTGAACATCTGTGACCGAGCGCGAGTACGCC 2160
QY 2167 CTGGGCACTATCCAGGCGCGAGCGCGCAAGAGCGAGCGAGCTGTGAAACCAATCATC 2226
DB 2161 CTGGGCACTATCCAGGCGCGAGCGCGCAAGAGCGAGCGAGCTGTGAAACCAATCATC 2220
QY 2227 GAGCGCTGATCAAAAGAGAGAGAGTGTACCTGAGACTGGTGGCCCGCCCAAGGGCATC 2286
DB 2221 GAGCGCTGATCAAAAGAGAGAGAGTGTACCTGAGACTGGTGGCCCGCCCAAGGGCATC 2280
QY 2287 GCGCGCAACGAGACATCGACAGCTGTGTAGAGCAAGGCGATCCGCAAGGTGCTGTCTG 2346
DB 2281 GCGCGCAACGAGACATCGACAGCTGTGTAGAGCAAGGCGATCCGCAAGGTGCTGTCTG 2340
QY 2347 GACGCGATCGATGCGGCGATCTGTATCTAACAGTACATGACGACCTGTATCTGTGGCAGC 2406
DB 2341 GACGCGATCGATGCGGCGATCTGTATCTAACAGTACATGACGACCTGTATCTGTGGCAGC 2400
QY 2407 GCGCGCGCTTGAATGATTTAAAGCTTCCCGGGGCTTACGACCGGCT 2451
DB 2401 GCGCGCGCTTGAATGATTTAAAGCTTCCCGGGGCTTACGACCGGCT 2445

RESULT 4
ADCl3264
ID ADCl3264 standard; DNA; 2445 BP.
AC ADCl3264;
XX
AC ADCl3264;
XX
DT 18-DEC-2003 (first entry)
XX
DB DNA of HIV construct p2Pol-opt-YMMW_C SEQ ID NO 43.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prox;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
OS Human immunodeficiency virus.
XX
PN MO2003004620-A2.
PD 16-JAN-2003.
PP 05-JUL-2002; 2002MO-US021420.
XX
XX 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megele J, Barnett SM, Llan Y, Engelbrecht S, Van Rensburg EJ;

XX
DR WPI, 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RNaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure: Fig 40; 301dp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RNaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;
Query Match 99.4%; Score 2443.4; DB 10; Length 2445;
Best Local Similarity 100.0%; Pred. No. 5.3e-294;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 7 GCCACATGCGCGAGGCGCATGACCGAGCGCAACATCTGTATGACGCGCAGC 66
DB 1 GCCACATGCGCGAGGCGCATGACCGAGCGCAACATCTGTATGACGCGCAGC 60
QY 67 AACTTCAAGGGCGCCCAAGCGATCATCAAGTCTTCACTCGCGAGAGAGGCGCACATC 126
DB 61 AACTTCAAGGGCGCCCAAGCGATCATCAAGTCTTCACTCGCGAGAGAGGCGCACATC 120
QY 127 GCCCGCAACTGCGCGCGCGCGCGCAAGAGAGGCTGCTGAAAGTGGCGAGAGGCGCAGC 186
DB 121 GCCCGCAACTGCGCGCGCGCGCGCGCAAGAGAGGCTGCTGAAAGTGGCGAGAGGCGCAGC 180
QY 187 CAGATGAAGAGCTGACCGAGCGCGCGCAACTTCTTCCGCGAGAGACTTGGCTTCCCG 246
DB 181 CAGATGAAGAGCTGACCGAGCGCGCGCAACTTCTTCCGCGAGAGACTTGGCTTCCCG 240
QY 247 CAGGCGAAGGCGCGCGGATTCGCCAGCGAGAGAGAGCGCGCAACGCGCGCAAGCGCGC 306
DB 241 CAGGCGAAGGCGCGCGGATTCGCCAGCGAGAGAGAGCGCGCGCAACGCGCGCAAGCGCGC 300
QY 307 GAGCTGAGTGGCGCGCGAGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGC 366
DB 301 GAGCTGAGTGGCGCGCGAGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGC 360
QY 367 CTGAACCTTCCCGCGAGTACCTGTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 426
DB 361 CTGAACCTTCCCGCGAGTACCTGTGGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 420
QY 427 CAGATGAAGAGAGCGCTGTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486
DB 421 CAGATGAAGAGAGCGCTGTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 480
QY 487 CTGCGCGCGAGTGAAGCGCGCAAGATGATCGCGCGCGCATCGCGCGCTTATCAAGGTGGCGC 546
DB 481 CTGCGCGCGAGTGAAGCGCGCAAGATGATCGCGCGCGCATCGCGCGCTTATCAAGGTGGCGC 540
QY 547 CAGTACGACAGATCTGTATGAGATCTGCGCGCAAGAGCGCGCATCGCGCGCTGTATC 606
DB 541 CAGTACGACAGATCTGTATGAGATCTGCGCGCAAGAGCGCGCATCGCGCGCTGTATC 600
QY 607 GCGCGCGCGCGCGCGAGATCATGCGCGCGCGCAAGTCTGAGACCGAGCTGGCTGCGCCTG 666
DB 601 GCGCGCGCGCGCGCGAGATCATGCGCGCGCGCAAGTCTGAGACCGAGCTGGCTGCGCCTG 660
QY 667 AACTTCCCATGAGCGCGCATGAGACGCTGCGCGGTGAAGTGAAGCGCGCGCATGAGCGCGC 726

```

Db      661  AACTTCCCATCAGCCCATCGAGACCGTGGCCCTGAAAGCTGAAGCCCGGCAATGACCGC 720
Qy      727  CCCAAGGTGAACAGTGGCCCTCTGAACGAGAGAAATCAAGAGCCCTGAACCGCATCTGC 786
Db      721  CCCAAGGTGAACAGTGGCCCTCTGAACGAGAGAAATCAAGAGCCCTGAACCGCATCTGC 780
Qy      787  GAGGAGATGAGAGAGAGAGGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTTACAACAC 846
Db      781  GAGGAGATGAGAGAGAGAGGCAAGATCACCAAGATCGGCCCCGAGAAACCCCTTACAACAC 840
Qy      847  CCCGGTTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
Db      841  CCCGGTTTCGCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy      907  GAGCTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 966
Db      901  GAGCTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
Qy      967  GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
Db      961  GAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Qy      1027  CCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
Db      1021  CCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Qy      1087  ACCCGGCGATCGCTACCAAGATCAACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
Db      1081  ACCCGGCGATCGCTACCAAGATCAACAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Qy      1147  ATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
Db      1141  ATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Qy      1207  GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
Db      1201  GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy      1267  AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
Db      1261  AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy      1327  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
Db      1321  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1387  CCCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
Db      1381  CCCATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy      1447  AAGCTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
Db      1441  AAGCTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy      1507  CTGCGGCGGCGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
Db      1501  CTGCGGCGGCGCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy      1567  CTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
Db      1561  CTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy      1627  AAGGAGCTGTGGCGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
Db      1621  AAGGAGCTGTGGCGAGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy      1687  CAGGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
Db      1681  CAGGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy      1747  AACGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806

```

```

Db      1741  AACGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy      1807  TGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
Db      1801  TGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy      1867  ACCGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
Db      1861  ACCGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy      1927  GTGAAGCTGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
Db      1921  GTGAAGCTGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy      1987  GACGCGCGCGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2046
Db      1981  GACGCGCGCGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy      2047  CGGCAAGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2106
Db      2041  CGGCAAGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy      2107  CAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2166
Db      2101  CAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy      2167  CTGGGCAATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226
Db      2161  CTGGGCAATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy      2227  GAGGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2286
Db      2221  GAGGAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy      2287  GGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2346
Db      2281  GGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy      2347  GAGGCAATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2406
Db      2341  GAGGCAATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy      2407  GGGGCGCTTGAAGATGATTTAAAGCTTCCGGGGGCTAAGACCGGT 2451
Db      2401  GGGGCGCTTGAAGATGATTTAAAGCTTCCGGGGGCTAAGACCGGT 2445

RESULT 5
ID      ADL13230
ID      ADL13230 standard; DNA; 3930 BP.
XX      ADL13230;
XX      18-DBC-2003 (first entry)
XX      ;
DE      DNA of HIV construct GagComp]Polmut_C SEQ ID NO 9.
XX      expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prox;
KW      Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
OS      Human immunodeficiency virus.
XX      WO2003004620-A2.
XX      PN
XX      PD
XX      16-JAN-2003.
XX      PF
XX      05-JUL-2002; 2002MO-US021420.
XX      PR
XX      05-JUL-2001; 2001US-0303192P.
XX      PR
XX      31-AUG-2001; 2001US-031680P.
XX      PR
XX      16-JAN-2002; 2002US-0349871P.
XX      PA
XX      (CHIR ) CHIRON CORP.

```



```

QY 1754 TGAAGCAGCTGACCGAGGCGCTGCAAGAGATCCCATGAGAGCATCTGTGATCTGGGGCA 1813
DB 3227 TGAAGCAGCTGACCGAGGCGCTGCAAGAGATCCCATGAGAGCATCTGTGATCTGGGGCA 3286
QY 1814 AGACCCCAAGTCTCGGCTGCTCCATCCAGAGAGAGACCTGGAGAGACTGGTGGAGCCGACT 1873
DB 3287 AGACCCCAAGTCTCGGCTGCTCCATCCAGAGAGAGACTGGTGGAGAGACTGGTGGAGCCGACT 3346
QY 1874 ACTGGCAGGCGCACCTGATCCCGAGTGGAGATTGCTGAACAACCCCGCTGTGTAAGC 1933
DB 3347 ACTGGCAGGCGCACCTGATCCCGAGTGGAGATTGCTGAACAACCCCGCTGTGTAAGC 3406
QY 1934 TGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1993
DB 3407 TGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3466
QY 1994 CCGCCACCGGAGAGACCAAGATCGGCAAGGCGGCTAGTGAACCGACCGGGGCGGCGAGA 2053
DB 3467 CCGCCACCGGAGAGACCAAGATCGGCAAGGCGGCTAGTGAACCGACCGGGGCGGCGAGA 3526
QY 2054 AGATGCTGAGCTGACCGAGACCAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2113
DB 3527 AGATGCTGAGCTGACCGAGACCAACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3586
QY 2114 CCTTGAAGAGACAGCGGACAGGAGTGAACATCTGTGAACCGACCGAGAGAGAGAGAGAGAG 2173
DB 3587 CCTTGAAGAGACAGCGGACAGGAGTGAACATCTGTGAACCGACCGAGAGAGAGAGAGAGAG 3646
QY 2174 TCATCAGAGCGGACCGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2233
DB 3647 TCATCAGAGCGGACCGGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3706
QY 2234 TGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2293
DB 3707 TGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3766
QY 2294 ACAGCAGATGCAACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
DB 3767 ACAGCAGATGCAACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3826
QY 2354 TCGATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2413
DB 3827 TCGATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3886
QY 2414 CTAGAGTCGATTTAAAGCTTCCCGGAGGCTAGACCGGCT 2451
DB 3887 CTAGAGTCGATTTAAAGCTTCCCGGAGGCTAGACCGGCT 3924

```

```

RESULT 6
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
AC ABL39960;
XX
XX 15-MAY-2002 (first entry)
DE Synthetic construct PR97SYM SEQ ID NO:31.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
XX MO200204493-A2.
XX
XX 17-JAN-2002.
XX
XX 05-JUL-2001; 2001MO-US021241.
XX

```

```

PR 05-JUL-2001; 2000US-00610313.
XX
XX (CHIR ) CHIRON CORP.
PA (UYST-) UNIV STILLENBOSCH.
XX
XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX WPI; 2002-154920/20.
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
XX Claim 1; Fig 9; 23pp; English.
XX
XX The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (1). (1) Have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;
XX
Query Match 99.2%; Score 2436.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 4,1e-293;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGCAAGCGCACCATTGAGCGGAGGCGCATGAGAGGCGCAGGCGGCGCAACATCTGATGAG 60
DB 1 GTGCAAGCGCACCATTGAGCGGAGGCGCATGAGAGGCGCAGGCGGCGCAACATCTGATGAG 60
QY 61 CGCAGCAACTTAAAGAGGCGGCGGCGCATGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 61 CGCAGCAACTTAAAGAGGCGGCGGCGCATGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 61 CGCAGCAACTTAAAGAGGCGGCGGCGCATGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
DB 61 CGCAGCAACTTAAAGAGGCGGCGGCGCATGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
QY 121 CACATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
DB 121 CACATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 181 GGCACACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 181 GGCACACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 TTCCCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
DB 241 TTCCCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 301 AGCCCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
DB 301 AGCCCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 361 GGCACCGTGAACCTTCCCGCAGATCAACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 361 GGCACCGTGAACCTTCCCGCAGATCAACCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
QY 421 GGCAGCGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
DB 421 GGCAGCGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
QY 481 ATGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
DB 481 ATGAGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 541 GTGCGCGAGTACAGACAGATCTGATGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
DB 541 GTGCGCGAGTACAGACAGATCTGATGAGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600

```

QY 601 CTGATGGGCCCCCCTGGTAAATCATGTGGGCGCAACATGTGACCCAGCTGGGCTGC 660
Db 601 CTGATGGGCCCCCCTGGTAAATCATGTGGGCGCAACATGTGACCCAGCTGGGCTGC 660
QY 661 ACCCTGAATCTTCCCATGAGGCGCCATGAGACCGTGGCCCGTGAAGCTGAAGCCCGGCAATG 720
Db 661 ACCCTGAATCTTCCCATGAGGCGCCATGAGACCGTGGCCCGTGAAGCTGAAGCCCGGCAATG 720
QY 721 GACGGCCCCAAGGTGAAGCATGTGGCCCTGACCGAGAGAAAGATCAAGAGCCCTGACCGCC 780
Db 721 GACGGCCCCAAGGTGAAGCATGTGGCCCTGACCGAGAGAAAGATCAAGAGCCCTGACCGCC 780
QY 781 ATCTCGAGAGATGAGAGAGAGGCAAGATCAACAAAGATGGCCCTCGAAGACCCCTAC 840
Db 781 ATCTCGAGAGATGAGAGAGAGGCAAGATCAACAAAGATGGCCCTCGAAGACCCCTAC 840
QY 841 AACACCCCGTGTTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db 841 AACACCCCGTGTTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCCGAGAGCTGAACAGAGGCAACCAAGGCTTCTGGAGGTGACCTGGGCAATCCCGAC 960
Db 901 TTCCGAGAGCTGAACAGAGGCAACCAAGGCTTCTGGAGGTGACCTGGGCAATCCCGAC 960
QY 961 CCGCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 961 CCGCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1021 AGCGTGGCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
Db 1021 AGCGTGGCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
Db 1081 AACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 CCGAGAGCTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
Db 1141 CCGAGAGCTTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Db 1201 GAGATGGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 CCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Db 1261 CCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
Db 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
QY 1375 TGGAGCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
Db 1375 TGGAGCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
QY 1435 AAGCTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
Db 1435 AAGCTGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
QY 1495 CTGTGCAAGCTGTGTCGCGCGCGCAAGGCGCTGACCGACATGTGTCGCGCGCGCAAGAG 1554
Db 1495 CTGTGCAAGCTGTGTCGCGCGCGCAAGGCGCTGACCGACATGTGTCGCGCGCGCAAGAG 1554
QY 1555 GCGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
Db 1555 GCGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
QY 1615 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
Db 1615 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
QY 1675 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
Db 1675 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734

Db 1681 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1735 ACCGCGCAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Db 1741 ACCGCGCAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1795 AGCATGTGATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Db 1801 AGCATGTGATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1855 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Db 1861 GAGAGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1915 ACCCGCGCGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
Db 1921 ACCCGCGCGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1975 ACCTTCTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
Db 1981 ACCTTCTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2035 ACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
Db 2041 ACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2095 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
Db 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2155 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
Db 2161 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2215 AACGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
Db 2221 AACGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2275 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
Db 2281 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2335 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2394
Db 2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
QY 2395 TACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2454
Db 2401 TACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
QY 2455 TTC 2457
Db 2461 TTC 2463

RESULT 7
ADM73765
ID ADM73765 standard; DNA; 2463 BP.
XX
XX ADM73765;
XX
XX 03-JUN-2004 (first entry)
XX
XX
XX HIV-1 polynucleotide #8.
XX
XX HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
XX
XX HIV type C protein; immunostimulant.
XX
XX Human immunodeficiency virus 1.
XX
XX US200323961-A1.

PD 04-DEC-2003.
 XX 05-JUL-2001; 2001US-00899575.
 XX 05-JUL-2000; 2000US-00610313.
 XX (MEGE/) MEGEDE J Z.
 PA (BARN/) BARNETT S W.
 PA (ENGE/) ENGELBRECHT S.
 PA (RENS/) RENSBURG E J V.
 XX Megede JZ, Barnett SM, Engelbrecht S, Rensburg EJV;
 XX WPI; 2004-060515/06.
 DR
 XX
 PT New expression cassette comprising a polynucleotide sequence encoding an
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
 PT immunisation, generating of packaging cell lines or in producing HIV Type
 PT C proteins.
 XX
 PS Claim 1; SEQ ID NO 31; 160pp; English.
 XX
 CC The invention relates to an expression cassette comprising a
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
 CC also relates to a recombinant expression system for use in a host cell
 CC comprising an expression cassette, where the polynucleotide sequence
 CC further comprises control elements capable of driving expression in the
 CC selected host cell, a cell comprising an expression cassette where the
 CC polynucleotide sequence further comprises control elements compatible
 CC with the expression in the cell and a composition for generating an
 CC immunological response, comprising an expression cassette. The expression
 CC cassette and the methods of the invention are useful in eliciting an
 CC immune response, in DNA immunisation, in generation of packaging cell
 CC lines and in producing HIV Type C proteins. This sequence represents an
 CC HIV-1 polynucleotide of the invention.
 XX
 SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;
 Query Match 99.2%; Score 2436.2; DB 12; Length 2463;
 Best Local Similarity 99.6%; Pred. No. 4.1e-293;
 Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 481 ATGAGCCGCGCCGCAAGTGAAGGCCCAAGTGAATCGGCGGCATCGGCGCTTCAACAG 540
 DB 481 ATGAGCCGCGCCGCAAGTGAAGGCCCAAGTGAATCGGCGGCATCGGCGCTTCAACAG 540
 QY 541 GTGCGCCAGTACGACCAATCCGATCGAGATCTGCGCGCAAGAGGCCATCGGCACTG 600
 DB 541 GTGCGCCAGTACGACCAATCCGATCGAGATCTGCGCGCAAGAGGCCATCGGCACTG 600
 QY 601 CTGATCGGCCCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGCGCTG 660
 DB 601 CTGATCGGCCCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGCGCTG 660
 QY 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCA 720
 DB 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGTGCCTCGTGAAGCTGAAGCCCGGCA 720
 QY 721 GACGCGCCCAAGGTGAACAGTGGCCCTGACCGAGAGAAAGATCAAGGCGCTGACCGCC 780
 DB 721 GACGCGCCCAAGGTGAACAGTGGCCCTGACCGAGAGAAAGATCAAGGCGCTGACCGCC 780
 QY 781 ATCTGCGAGAGATGAGAGAGAGGCAAGATCAACCAAGATCGGCCCGAGAACCCCTAC 840
 DB 781 ATCTGCGAGAGATGAGAGAGAGGCAAGATCAACCAAGATCGGCCCGAGAACCCCTAC 840
 QY 841 AACACCCCGCTTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 DB 841 AACACCCCGCTTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 901 TTCGCGAGCTGAACAGAGCGCAACCGAGACTTCTGAGAGTGCAGCTGAGGATCCCGC 960
 DB 901 TTCGCGAGCTGAACAGAGCGCAACCGAGACTTCTGAGAGTGCAGCTGAGGATCCCGC 960
 QY 961 CCGCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 DB 961 CCGCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1021 AGCGTCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 DB 1021 AGCGTCCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 QY 1081 AACGAGACCCCGGAGATCGGCTTCAACAGATCTGAGAGAGAGAGAGAGAGAGAGAG 1140
 DB 1081 AACGAGACCCCGGAGATCGGCTTCAACAGATCTGAGAGAGAGAGAGAGAGAGAGAG 1140
 QY 1141 CCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 DB 1141 CCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1201 GAGATCGTGAATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 DB 1201 GAGATCGTGAATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1261 GCGGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 DB 1261 GCGGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
 QY 1321 AAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
 DB 1321 AAGAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
 QY 1375 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
 DB 1375 TGAACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
 QY 1435 AAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
 DB 1435 AAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
 QY 1495 CTGTGCAAGCTCTGTGCGCGCGCAAGGCGCTGACCGGATCTGTGCGCTGACCGAGAG 1554
 DB 1495 CTGTGCAAGCTCTGTGCGCGCGCAAGGCGCTGACCGGATCTGTGCGCTGACCGAGAG 1554
 QY 1501 CTGTGCAAGCTCTGTGCGCGCGCAAGGCGCTGACCGGATCTGTGCGCTGACCGAGAG 1560
 DB 1501 CTGTGCAAGCTCTGTGCGCGCGCAAGGCGCTGACCGGATCTGTGCGCTGACCGAGAG 1560

```

OY 1555 GCCGAGCTGAGCTGAGCGAGAACCGGAGATCTCTGCGGAGCCCGTCACGCGCTGTAC 1614
DB 1561 GCCGAGCTGAGCTGAGCGAGAACCGGAGATCTCTGCGGAGCCCGTCACGCGCTGTAC 1620
OY 1615 TACGACCCCAAGAGAGAGCTGTGTGCGGAGATCTCAAGAGCGAGCGACGACGATGTAG 1674
DB 1621 TACGACCCCAAGAGAGAGCTGTGTGCGGAGATCTCAAGAGCGAGCGACGACGATGTAG 1680
OY 1675 TACGAGATCTACGAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAAGCCAGATGAGC 1734
DB 1681 TACGAGATCTACGAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTAAGCCAGATGAGC 1740
OY 1735 ACCGCCCAACCAACGAGCTGAGAGAGCTGACCGAGCGCTGTGACAGAGATGCGCATGAG 1794
DB 1741 ACCGCCCAACCAACGAGCTGAGAGAGCTGACCGAGCGCTGTGACAGAGATGCGCATGAG 1800
OY 1795 AGCATGTGATCTGAGGAGAGAGCCCGCAAGTTCCCGCTGCGCATTCAGAGAGAGAGCTGAG 1854
DB 1801 AGCATGTGATCTGAGGAGAGAGCCCGCAAGTTCCCGCTGCGCATTCAGAGAGAGAGCTGAG 1860
OY 1855 GAGACCTGTGAGAGCGAGCTACTGAGAGAGCGACCTGATCCCGAGTGGAGATTGCTGAGAC 1914
DB 1861 GAGACCTGTGAGAGCGAGCTACTGAGAGAGCGACCTGATCCCGAGTGGAGATTGCTGAGAC 1920
OY 1915 ACCCCCCCTGTGTGAGAGCTGTGTACAGAGCTGAGAGAGAGCCCATCATGCGCGCGAG 1974
DB 1921 ACCCCCCCTGTGTGAGAGCTGTGTGTACAGAGCTGAGAGAGAGCCCATCATGCGCGCGAG 1980
OY 1975 ACCTTCTACGTGAGAGAGAGCGCGCGACACCGGACGACAGATGCGGCAAGCGCGCTAGCTG 2034
DB 1981 ACCTTCTACGTGAGAGAGAGCGCGCGCGACACCGGACGACAGATGCGGCAAGCGCGCTAGCTG 2040
OY 2035 ACCGACCGGAGCGCGGAGAGAGTGTGAGAGCTGACCGAGACCGACCAACGAGAGAGCGAG 2094
DB 2041 ACCGACCGGAGCGCGGAGAGAGTGTGAGAGCTGACCGAGACCGACCAACGAGAGAGCGAG 2100
OY 2095 CTGACAGGCGATCTCAAGCTGCGCTGTGAGAGACGCGGACGAGAGTGAATCGTGTGACGAC 2154
DB 2101 CTGACAGGCGATCTCAAGCTGCGCTGTGAGAGACGCGGACGAGAGTGAATCGTGTGACGAC 2160
OY 2155 AGCCAGTACGCGCTGTGAGAGCTCAACAGGCGCGACCGGACGAGAGAGAGAGAGAGCTGTG 2214
DB 2161 AGCCAGTACGCGCTGTGAGAGCTCAACAGGCGCGACCGGACGAGAGAGAGAGAGAGCTGTG 2220
OY 2215 AACGAGATCATGAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGCGTGCCTGCGC 2274
DB 2221 AACGAGATCATGAGAGAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGCGTGCCTGCGC 2280
OY 2275 CACAGAGGCGATCGCGGAGACGAGAGATCGACCAAGCTGTGTGAGAGAGAGGCGATCGGAC 2334
DB 2281 CACAGAGGCGATCGCGGAGACGAGAGATCGACCAAGCTGTGTGAGAGAGAGGCGATCGGAC 2340
OY 2335 GTGCTGTCTGTGAGAGAGAGATGAGCGGCGATGTGTATCTACAGTATCATGAGAGAGCTG 2394
DB 2341 GTGCTGTCTGTGAGAGAGATGAGCGGCGATGTGTATCTACAGTATCATGAGAGAGCTG 2400
OY 2395 TACGAGGAGAGAGAGCGCGCTTGTAGATCGATTAAAGCTTCCGCGGCGTAAAGAGCGGTGA 2454
DB 2401 TACGAGGAGAGAGAGCGCGCTTGTAGATCGATTAAAGCTTCCGCGGCGTAAAGAGCGGTGA 2460
OY 2455 TTC 2457
DB 2461 TTC 2463

```

RESULT 8
ADCI3231
ID ADCI3231 standard; DNA; 3930 BP.

XX AC ADCI3231;
XX DT 18-DEC-2003 (first entry)
XX

```

DE DNA of HIV construct GagComp1Polmutat_C SHQ ID NO 10.
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; ProC;
KM Rev; HIV type C; gene therapy; vaccine; immunization; HIV; ds.
XX Human immunodeficiency virus.
XX WO200304620-A2.
XX 16-JAN-2003.
XX 05-JUN-2002; 2002WO-US021420.
XX 05-JUN-2001; 2001US-0303192P.
XX 31-AUG-2001; 2001US-0316860P.
XX 16-JAN-2002; 2002US-0349871P.
XX (CHIR ) CHIRON CORP.
XX (UYST-) UNIV STRLEMBOSCH.
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ,
XX WPI, 2003-221593/21.
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,
XX ProC, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX Disclosure; Fig 7; 301pp; English.
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RaseH, Pol, Tat, ProC, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;
Query Match 99.1%; Score 2434.8; DB 10; Length 3930;
Best Local Similarity 99.9%; Pred.No. 5.7e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 14 TGCCGAGGCGCATGAGCGAGCGACGACGCGCAATCTGTATGAGCGAGCAATTTCA 73
DB 1487 TGCCGAGGCGCATGAGCGAGCGACGACGCGCAATCTGTATGAGCGAGCAATTTCA 1546
OY 74 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATCGCGGACGAGAGGCGCATTCGCGCGCA 133
DB 1547 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATCGCGGACGAGAGGCGCATTCGCGCGCA 1606
OY 134 ACTGCGCGCGCGCGCGCGCAAGAGAGGCTGTGAGAGTGTGCGGACGAGAGGCGCACGAGATGA 193
DB 1607 ACTGCGCGCGCGCGCGCGCGCAAGAGAGGCTGTGAGAGTGTGCGGACGAGAGGCGCACGAGATGA 1666
OY 194 AGGACTGACCGAGCGCGCGCGCAATCTTCCGCGGAGGACCTGAGCTTCCCGCAGGCGCA 253
DB 1667 AGGACTGACCGAGCGCGCGCGCAATCTTCCGCGGAGGACCTGAGCTTCCCGCAGGCGCA 1726
OY 254 AGGCGCGGAGTTCCCGCAGCGAGAGAACCGCGCAACGCGCGCACGAGCGGAGCTGC 313
DB 1727 AGGCGCGGAGTTCCCGCAGCGAGAGAACCGCGCAACGCGCGCACGAGCGGAGCTGC 1786
OY 314 AGGTGCGCGGAGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCGAGGCGCACTTGAATCT 373
DB 1787 AGGTGCGCGGAGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCGAGGCGCACTTGAATCT 1846

```

QY 374 TCCCCAGATGACCTGTTGGCAGCCGCTTGTGAGATCAAGGTGGCGGCAGATCA 433
| | | | |
Db 1847 TCCCCAGATGACCTGTTGGCAGCCGCTTGTGAGATCAAGGTGGCGGCAGATCA 1906
| | | | |
QY 434 AGAGAGCCCTGTGGAACACCGGCGCGAGACACCGTGTGAGAGATGAGCTTGGCCG 493
| | | | |
Db 1907 AGAGAGCCCTGTGGAACACCGGCGCGAGACACCGTGTGAGAGATGAGCTTGGCCG 1966
| | | | |
QY 494 GCAGGTGAGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGCATGAG 553
| | | | |
Db 1967 GCAGGTGAGAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGGCGCATGAG 2026
| | | | |
QY 554 ACCAATCTGATTCAGATCTGCGGAGAGAGAGCCATCGGCAACCGTGTGATTCGGCCCA 613
| | | | |
Db 2027 ACCAATCTGATTCAGATCTGCGGAGAGAGAGCCATCGGCAACCGTGTGATTCGGCCCA 2086
| | | | |
QY 614 CCCCCGTGACATCATCTGCGCGCAGACATCTGACCCAGTGGAGTTCACCTGAACTTCC 673
| | | | |
Db 2087 CCCCCGTGACATCATCTGCGCGCAGACATCTGACCCAGTGGAGTTCACCTGAACTTCC 2146
| | | | |
QY 674 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGAGCCCGCATGAGACGGCCCCAAG 733
| | | | |
Db 2147 CCATCAGCCCATCGAGACCGTGCCTGTGAGCTGAGAGCCCGCATGAGACGGCCCCAAG 2206
| | | | |
QY 734 TGAAGCATGTGCCCCCTGACCGAGAGAGAGATCAAGGCCCTTGAACCGCATCTGCGAGAGAG 793
| | | | |
Db 2207 TGAAGCATGTGCCCCCTGACCGAGAGAGAGATCAAGGCCCTTGAACCGCATCTGCGAGAGAG 2266
| | | | |
QY 794 TGGAGAGAGAGGGCAAGATCAACAAGATGGGCCCGAGAGACCCCTTCAACAACCCCGGTGT 853
| | | | |
Db 2267 TGGAGAGAGAGGGCAAGATCAACAAGATGGGCCCGAGAGACCCCTTCAACAACCCCGGTGT 2326
| | | | |
QY 854 TCGCATCAAGAGAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGAGACTTCCGAGAGCTGA 913
| | | | |
Db 2327 TCGCATCAAGAGAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGAGACTTCCGAGAGCTGA 2386
| | | | |
QY 914 ACAAGGCAACCCAGAGACTTCTGGAGAGTGAAGTGGGCAATCCCCACCCCGCGGCTTGA 973
| | | | |
Db 2387 ACAAGGCAACCCAGAGACTTCTGGAGAGTGAAGTGGGCAATCCCCACCCCGCGGCTTGA 2446
| | | | |
QY 974 AAGAGAGAGAGAGCGGAGACGATGCTGAGAGTGGGCGAACGCTTCAACGATGGGCCCTTGA 1033
| | | | |
Db 2447 AAGAGAGAGAGAGCGGAGACGATGCTGAGAGTGGGCGAACGCTTCAACGATGGGCCCTTGA 2506
| | | | |
QY 1034 ACAGAGACTTCCGAGAGTACACCGGCTTCAACATCCCGAGCATCAACAAGAGACCCCG 1093
| | | | |
Db 2507 ACAGAGACTTCCGAGAGTACACCGGCTTCAACATCCCGAGCATCAACAAGAGACCCCG 2566
| | | | |
QY 1094 GCATCCGCTACCAAGTACAAAGTGTCTCCCAAGGCTTGAAGAGGAGAGCCCAAGCATTTCC 1153
| | | | |
Db 2567 GCATCCGCTACCAAGTACAAAGTGTCTCCCAAGGCTTGAAGAGGAGAGCCCAAGCATTTCC 2626
| | | | |
QY 1154 AAGAGCAGCATGACCAAGATCTGTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATGTGTATCT 1213
| | | | |
Db 2627 AAGAGCAGCATGACCAAGATCTGTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATGTGTATCT 2686
| | | | |
QY 1214 ACCAGGCCCCCGCTGATGTTGGGAGAGACCTGAGATTCGGCCAGCAACCGGCGCAAGATG 1273
| | | | |
Db 2687 ACCAGGCCCCCGCTGATGTTGGGAGAGACCTGAGATTCGGCCAGCAACCGGCGCAAGATG 2746
| | | | |
QY 1274 AGAGAGTGGCGCAAGCATCTGTGCTGCTGGGCTTCAACAACCCCGCAACAAGAGACACAGA 1333
| | | | |
Db 2747 AGAGAGTGGCGCAAGCATCTGTGCTGCTGGGCTTCAACAACCCCGCAACAAGAGACACAGA 2806
| | | | |
QY 1334 AGAGAGCCCCCTTCTGTGCTGCTGATGAGTGTGACCCCGCAAGATGTGACCCGATCTG 1393
| | | | |
Db 2807 AGAGAGCCCCCTTCTGTGCTGCTGATGAGTGTGACCCCGCAAGATGTGACCCGATCTG 2866
| | | | |
QY 1394 AGCTGCCGAGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGTGTGTGGGCAAGCTGA 1453
| | | | |
Db 2867 AGCTGCCGAGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGTGTGTGGGCAAGCTGA 2926
| | | | |
QY 1454 ACTGGGCGACAGATCTACCCCGGCAATCAAGGTGGCGCAGCTGTGCAAGTCTGTGCGG 1513
| | | | |

Db 2927 ACTGGGCGACAGATCTACCCCGGCAATCAAGTGTGCGCAAGTGTGCAAGCTGTGCGG 2986
| | | | |
QY 1514 GCGCCAGAGCCCTGACCGACATCTGTGCCCTTGAACGAGAGAGCGGAGCTGAGAGTGGCCG 1573
| | | | |
Db 2987 GCGCCAGAGCCCTGACCGACATCTGTGCCCTTGAACGAGAGAGCGGAGCTGAGAGTGGCCG 3046
| | | | |
QY 1574 AGAATCCGAGATCTGTGGCGGAGCCCGTGTGCAAGGCTGTATCTACAGACCCCAAGAGAAC 1633
| | | | |
Db 3047 AGAATCCGAGATCTGTGGCGGAGCCCGTGTGCAAGGCTGTATCTACAGACCCCAAGAGAAC 3106
| | | | |
QY 1634 TGGTGGCGAGATTCAGAGAGAGGCGCAAGACCAATGAGACTTACCAAGATCTACAGAGAC 1693
| | | | |
Db 3107 TGGTGGCGAGATTCAGAGAGAGGCGCAAGACCAATGAGACTTACCAAGATCTACAGAGAC 3166
| | | | |
QY 1694 CTTTCAAGACCTGAAAGACCGGCAAGTACCGCAAGATGCGCACCGGCCACCAAGAGAG 1753
| | | | |
Db 3167 CTTTCAAGACCTGAAAGACCGGCAAGTACCGCAAGATGCGCACCGGCCACCAAGAGAG 3226
| | | | |
QY 1754 TGAACAGCTGACCGAGGCGGTGTGAGAGATTCGCAATGAGAGCATCTGTATCTTGGGCA 1813
| | | | |
Db 3227 TGAACAGCTGACCGAGGCGGTGTGAGAGATTCGCAATGAGAGCATCTGTATCTTGGGCA 3286
| | | | |
QY 1814 AGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGAGACCTGGGAGACTGTGTGAGACGACT 1873
| | | | |
Db 3287 AGACCCCAAGTTCGCGCTGCGCCATCCAGAGAGAGACCTGGGAGACTGTGTGAGACGACT 3346
| | | | |
QY 1874 ACTGGCAGGCGACCTGATTCGCCGAGTGGAGTTCGTGAACAACCCCGCTGTGTGAGAC 1933
| | | | |
Db 3347 ACTGGCAGGCGACCTGATTCGCCGAGTGGAGTTCGTGAACAACCCCGCTGTGTGAGAC 3406
| | | | |
QY 1934 TGTGTATCAAGCTGAGAGAGAGAGCCCATCATCTGCGCGCGAGACTTCTACGTGAGACGGCG 1993
| | | | |
Db 3407 TGTGTATCAAGCTGAGAGAGAGAGCCCATCATCTGCGCGCGAGACTTCTACGTGAGACGGCG 3466
| | | | |
QY 1994 CCGCCACCGCGAGACCAAGATTCGGCAGAGCCGCGCTGATGAGACCGAGCGGCGCGGAG 2053
| | | | |
Db 3467 CCGCCACCGCGAGACCAAGATTCGGCAGAGCCGCGCTGATGAGACCGAGCGGCGCGGAG 3526
| | | | |
QY 2054 AGATGTGAGCTTGAACCGAGACCAACAACCAAGAGACCGAGCTGACAGCTGACAGCTG 2113
| | | | |
Db 3527 AGATGTGAGCTTGAACCGAGACCAACAACCAAGAGACCGAGCTGACAGCTGACAGCTG 3586
| | | | |
QY 2114 CCTTCAAGGACAGCGGACGAGGAGTGAACATCTGTGACCGACAGCGATACGCTTGGCA 2173
| | | | |
Db 3587 CCTTCAAGGACAGCGGACGAGGAGTGAACATCTGTGACCGACAGCGATACGCTTGGCA 3646
| | | | |
QY 2174 TCATCAAGGCGCAGGCGCCAGCAAGAGCGAGAGGAGCTGTGAAACCAAGATCATCGAGCAGC 2233
| | | | |
Db 3647 TCATCAAGGCGCAGGCGCCAGCAAGAGCGAGAGGAGCTGTGAAACCAAGATCATCGAGCAGC 3706
| | | | |
QY 2234 TGATCAAGAGAGAGAGTGTACTGAGCTGTGGGTGCCGCCCAAGAGGCAATCGGCGGCA 2293
| | | | |
Db 3707 TGATCAAGAGAGAGAGTGTACTGAGCTGTGGGTGCCGCCCAAGAGGCAATCGGCGGCA 3766
| | | | |
QY 2294 ACAGAGAGATTCAGCAAGGCTGTGAGCAAGGAGCATTCGCAAGGTGCTTCTTGAACGCA 2353
| | | | |
Db 3767 ACAGAGAGATTCAGCAAGGCTGTGAGCAAGGAGCATTCGCAAGGTGCTTCTTGAACGCA 3826
| | | | |
QY 2354 TCGATGGCGGCAATCGTATCTACAGTACATGAGAGACTTGTACGTTGGGCAAGCGGCGC 2413
| | | | |
Db 3827 TCGATGGCGGCAATCGTATCTACAGTACATGAGAGACTTGTACGTTGGGCAAGCGGCGC 3886
| | | | |
QY 2414 CTAGAGTGAATTAAGCTTCCCGGAGGCTAGACCGGT 2451
| | | | |
Db 3887 CTAGAGTGAATTAAGCTTCCCGGAGGCTAGACCGGT 3924
| | | | |

RESULT 9
ADCI3232
ID ADCI3232 standard; DNA; 3930 BP.
XX
AC ADCI3232;


```

Db 2867 AGCTCCCGAGAGAGAGAGCTGAGCCGTGAACGATCCAGAGAGCTGTGGCGAAGCTGA 2926
QY 1454 ACTGGGCGCAGCAGATCTAACCCTGGCATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCG 1513
Db 2927 ACTGGGCGCAGCAGATCTAACCCTGGCATCAAGGTGCGCGCCAGCTGTGCAAGCTGTGCGCG 2986
QY 1514 GCGCCAGAGGCGCTGACCGGACATCGTGTCCCTGACCGAGAGAGGCGCGAGCTGTGAGCTGAGCGG 1573
Db 2987 GCGCCAGAGGCGCTGACCGGACATCGTGTCCCTGACCGAGAGAGGCGCGAGCTGTGAGCTGAGCGG 3046
QY 1574 AGAAGCGCAGAGATCTGCGCGAGCCGCTGACAGCGCGTGTACTACGACCCCAAGAGAAC 1633
Db 3047 AGAAGCGCAGAGATCTGCGCGAGCCGCTGACAGCGCGTGTACTACGACCCCAAGAGAAC 3106
QY 1634 TGGTGGCCGAGATCCAGAGAGAGGCGCAGCAGCAGTGAAGCTTACCAAGATCTACAGAGAC 1693
Db 3107 TGGTGGCCGAGATCCAGAGAGAGGCGCAGCAGCAGTGAAGCTTACCAAGATCTACAGAGAC 3166
QY 1694 CCTTCAGAGAACTGTAGAGAGCCGCGCAGTACGCCCAAGATGCGCACCCGCGCACACCAAGAG 1753
Db 3167 CCTTCAGAGAACTGTAGAGAGCCGCGCAGTACGCCCAAGATGCGCACCCGCGCACACCAAGAG 3226
QY 1754 TGAAGCAGCTGACCGAGGCGCTGACAGAGATGCGCATGAGAGCAGTCTGTATCTGAGGAG 1813
Db 3227 TGAAGCAGCTGACCGAGGCGCTGACAGAGATGCGCATGAGAGCAGTCTGTATCTGAGGAG 3286
QY 1814 AGACCCCGCAAGTTCGCGCTGCGCATTCAGAGAGAGACCTTGGAGAGCTGTGTGAGCCGACT 1873
Db 3287 AGACCCCGCAAGTTCGCGCTGCGCATTCAGAGAGAGACCTTGGAGAGCTGTGTGAGCCGACT 3346
QY 1874 ACTGGCAGGCGCAGCTGTGATCCCGAGATGCGAGATTCGTGAACACCCCGCGCTGTGTGAAC 1933
Db 3347 ACTGGCAGGCGCAGCTGTGATCCCGAGATGCGAGATTCGTGAACACCCCGCGCTGTGTGAAC 3406
QY 1934 TGGGTGACCACTGAGAGAGAGAGCCCATCATCGGCGCGCAGACCTTCTTACGTGTGAGCGG 1993
Db 3407 TGGGTGACCACTGAGAGAGAGAGCCCATCATCGGCGCGCAGACCTTCTTACGTGTGAGCGG 3466
QY 1994 CCGCCAGCCGAGAGCAAGATTCGAGAGAGCCGCGCTACGTGACCGACCGGCGCGCGCAGA 2053
Db 3467 CCGCCAGCCGAGAGCAAGATTCGAGAGAGCCGCGCTACGTGACCGACCGGCGCGCGCAGA 3526
QY 2054 AGATCTGAGCTGACCGAGAGACCAACCAAGAGAGCCGAGCTGTGAGCCATCAGCTGTG 2113
Db 3527 AGATCTGAGCTGACCGAGAGACCAACCAAGAGAGCCGAGCTGTGAGCCATCAGCTGTG 3586
QY 2114 CCTGTGAGAGAGAGGCGCAGCGAGTGAACATCGTGAACCGACGCGCTGAGCCCTGTGGGA 2173
Db 3587 CCTGTGAGAGAGAGGCGCAGCGAGTGAACATCGTGAACCGACGCGCTGAGCCCTGTGGGA 3646
QY 2174 TCATTCAGGCGCCAGCCGACAGAGAGAGAGCGAGCTGTGTAACAGATCATGACACAGC 2233
Db 3647 TCATTCAGGCGCCAGCCGACAGAGAGAGAGCGAGCTGTGTAACAGATCATGACACAGC 3706
QY 2234 TGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCCGCAAGAGGCGATCGCGGCA 2293
Db 3707 TGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCCGCAAGAGGCGATCGCGGCA 3766
QY 2294 AGAGAGAGATTCGACAGCTGTGTGAGAGAGGCGATTCGCGCAAGGTGTCTTGTGAGAGGGA 2353
Db 3767 AGAGAGAGATTCGACAGCTGTGTGAGAGAGGCGATTCGCGCAAGGTGTCTTGTGAGAGGGA 3826
QY 2354 TCAGTGGCGGCACTGTGATCTACAGATCAATGAGACGACTGTAGTGGCGAGCGCGCC 2413
Db 3827 TCAGTGGCGGCACTGTGATCTACAGATCAATGAGACGACTGTAGTGGCGAGCGCGCC 3886
QY 2414 CTAGATTCGATTAAAGCTTCCCGGCGCTAGCACCGGT 2451
Db 3887 CTAGATTCGATTAAAGCTTCCCGGCGCTAGCACCGGT 3924

```

RESULT 10
ACA03591

```

ID ACA03591 standard; DNA; 5184 BP.
XX
AC ACA03591;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #74.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PP 05-JUL-2002; 2002MO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Zur Megele J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 79; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SO Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;

Query Match 99.1%; Score 2434.8; DB 8; Length 5184;
Best Local Similarity 99.9%; Pred. No. 5; 5e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 14 TGGCGGAGGCGCATGAGCGCAGGCGCAGGCGCAACATCTGTATGAGCGCGCAACTTCA 73
Db 2741 TGGCGGAGGCGCATGAGCGCAGGCGCAGGCGCAGGCGCAACATCTGTATGAGCGCGCAACTTCA 2800
QY 74 AGGCGCCCAAGGCGCATCATCAAGTGTCTTCACTGCGGCAAGAGAGGCGCATGCGCCGCA 133
Db 2801 AGGCGCCCAAGGCGCATCATCAAGTGTCTTCACTGCGGCAAGAGAGGCGCATGCGCCGCA 2860
QY 134 ACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193
Db 2861 ACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2920
QY 194 AGGACTGACCGAGCGCGCAGGCGCAACTTCTTCCGCGAGAGACTGTGCGCTTCCCGAGGGA 253
Db 2921 AGGACTGACCGAGCGCGCAGGCGCAACTTCTTCCGCGAGAGACTGTGCGCTTCCCGAGGGA 2980

```

254 AGGCGCGAGTTCCCGACGAGCAGAAACCGGCGCAAGGCCGACGACGCGGAGCTGC 313
Db AGGCGCGAGTTCCCGACGAGCAGAAACCGGCGCAAGGCCGACGACGCGGAGCTGC 3040
Qy AGGTGCGGCGCAGAACCCCGCAGCGAGGCGCGGCGCGAGGCGCAGGCGCAGCTGAACT 373
Db AGGTGCGGCGCAGAACCCCGCAGCGAGGCGCGGCGCGAGGCGCAGGCGCAGCTGAACT 3100
Qy TCCCCAATATCACTCTGTGGCAGGCGCCCTGTGTAGATCAAGGTGGCGGCGCAGATCA 433
Db TCCCCAATATCACTCTGTGGCAGGCGCCCTGTGTAGATCAAGGTGGCGGCGCAGATCA 3160
Qy AGGAGGCGCTGTGGCAGCGGCGCGCAGCAGACCGTGTGGAGAGATGAGCCCTGCGCG 493
Db AGGAGGCGCTGTGGCAGCGGCGCGCAGCAGACCGTGTGGAGAGATGAGCCCTGCGCG 3220
Qy GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGCGCCAGTACG 553
Db GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGCGCCAGTACG 3280
Qy ACCAGATCTGTATCGAGATCTGTGGCGAAGAGGCGCATGGGCAACCGTGTGATCGGCGCA 613
Db ACCAGATCTGTATCGAGATCTGTGGCGAAGAGGCGCATGGGCAACCGTGTGATCGGCGCA 3340
Qy CCCCCGTGAACATCATCGGCGCGCAATGTCTGACCGAGTGGGCTGCAACCTGAACTTCC 673
Db CCCCCGTGAACATCATCGGCGCGCAATGTCTGACCGAGTGGGCTGCAACCTGAACTTCC 3400
Qy CCATCAGCCCATTCAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGAGACGCGCCCAAG 733
Db CCATCAGCCCATTCAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGAGACGCGCCCAAG 3460
Qy TGAACAGTGGGCGCTGTGACCGAGAGAGATCAAGGCGCTGACCGGCAATCTGTGAGAGAG 793
Db TGAACAGTGGGCGCTGTGACCGAGAGAGATCAAGGCGCTGACCGGCAATCTGTGAGAGAG 3461
Qy TGAACAGTGGGCGCTGTGACCGAGAGAGATCAAGGCGCTGACCGGCAATCTGTGAGAGAG 3520
Db TGAACAGTGGGCGCTGTGACCGAGAGAGATCAAGGCGCTGACCGGCAATCTGTGAGAGAG 794
Qy TGAAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 853
Db TGAAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 3521
Qy TGGGAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 913
Db TGGGAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 3581
Qy TGGGAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 914
Db TGGGAGAGAGGCGCAAGATCAACAAGTGGGCGCGGAGAACCCCTTCAACAACCCCGGAGT 3641
Qy ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCAACCCCGCGGCGTGA 973
Db ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCAACCCCGCGGCGTGA 3641
Qy ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCAACCCCGCGGCGTGA 974
Db ACAAGCGCACCCAGGACTTCTGGAGGTGCAAGTGGGCAATCCCAACCCCGCGGCGTGA 3701
Qy AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 1033
Db AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 3701
Qy AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 1034
Db AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 3761
Qy AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 1094
Db AGAAGAGAGAGGCGTGAACCGTGTGGAGCGTGGGCGAGCGCTTCAAGCGCGTGGCGCTGG 3821
Qy GCATCGGCTACCAAGTCAACGCTGCGCCAGAGGCTGGAAGGCGAGCGCCAGCATCTTCC 1153
Db GCATCGGCTACCAAGTCAACGCTGCGCCAGAGGCTGGAAGGCGAGCGCCAGCATCTTCC 3821
Qy AGAAGAGAGAGGCGTGAACCGTGTGGAGCGCTTCAAGCGCGTGGCGCTGGAGTCT 1213
Db AGAAGAGAGAGGCGTGAACCGTGTGGAGCGCTTCAAGCGCGTGGCGCTGGAGTCT 3881
Qy ACCAGGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 1273
Db ACCAGGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 3941
Qy ACCAGGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 1274
Db ACCAGGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 4001
Qy AGAAGCTGCGCAAGGACTGTGCGCTGGGCGCTTCAACAACCCCGCAAGAGAGCGCAGA 1333
Db AGAAGCTGCGCAAGGACTGTGCGCTGGGCGCTTCAACAACCCCGCAAGAGAGCGCAGA 4060

Qy AGAAGCGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 1334
Db AGAAGCGCGCGCTGTACGTTGCGCAGCGACCTGTGAGATCGGCGCAAGGCGCAAGATCG 4061
Qy AGTTCGCGAGAGAGAGGCTTGAACCGTGAACGATTCAGAACTGTGTGGGCAAGCTGA 1453
Db AGTTCGCGAGAGAGAGGCTTGAACCGTGAACGATTCAGAACTGTGTGGGCAAGCTGA 4121
Qy ACTGGGCGAGGAGATCTAACCGCGGAGCTGAAGGTGGCGGAGGCTGTGGCGG 1513
Db ACTGGGCGAGGAGATCTAACCGCGGAGCTGAAGGTGGCGGAGGCTGTGGCGG 4181
Qy GCGCAAGGCGCTTGAACGATCGTGCCTGACCGAGAGAGGCGAGCTGAGCTGCGCG 1573
Db GCGCAAGGCGCTTGAACGATCGTGCCTGACCGAGAGAGGCGAGCTGAGCTGCGCG 4241
Qy AGAAGCGGAGATCTGTGCGAGGCGCGTGAACCGGCGTGTACTAGACCCGAGCAAGAGCC 1633
Db AGAAGCGGAGATCTGTGCGAGGCGCGTGAACCGGCGTGTACTAGACCCGAGCAAGAGCC 4301
Qy TGTGGCGGAGATCTCAAGAGCGAGGCGCAGACAGTGAACCTTACCAATCTTACAGAGAGC 1693
Db TGTGGCGGAGATCTCAAGAGCGAGGCGCAGACAGTGAACCTTACCAATCTTACAGAGAGC 4361
Qy CTTCAAGAACTTGAAGACGCGCAAGTACGCGCAAGATGCGACCGGCGCAACAGAG 1753
Db CTTCAAGAACTTGAAGACGCGCAAGTACGCGCAAGATGCGACCGGCGCAACAGAG 4421
Qy TGAAGCAGCTGACCGAGGCGGTGCAAGATCGCATGAGAGCATCTGTGATCTTGGGCA 1813
Db TGAAGCAGCTGACCGAGGCGGTGCAAGATCGCATGAGAGCATCTGTGATCTTGGGCA 4481
Qy TGAAGCAGCTGACCGAGGCGGTGCAAGATCGCATGAGAGCATCTGTGATCTTGGGCA 1814
Db TGAAGCAGCTGACCGAGGCGGTGCAAGATCGCATGAGAGCATCTGTGATCTTGGGCA 4541
Qy AGAAGCGGAGATCTGAGGCGGCGTGAACCGAGAGAGAGCTTGAAGAGCTGTGAGAGAG 1874
Db AGAAGCGGAGATCTGAGGCGGCGTGAACCGAGAGAGAGCTTGAAGAGCTGTGAGAGAG 4601
Qy ACTGGCAGGCGCACTGATGCCGAGTGGAGGTTGTGAACACCCCGCGCTGTGAGAGC 1933
Db ACTGGCAGGCGCACTGATGCCGAGTGGAGGTTGTGAACACCCCGCGCTGTGAGAGC 4601
Qy TGTGGTACCACTGAGAGAGAGGCGCATCATCGGCGCGAGACCTTCTACGTGAGAGCG 1993
Db TGTGGTACCACTGAGAGAGAGGCGCATCATCGGCGCGAGACCTTCTACGTGAGAGCG 4661
Qy CCGGCAACCGGAGACCAAGATCGGCAAGGCGCGGCTTACGTGAACCGGCGCGGCGAG 2053
Db CCGGCAACCGGAGACCAAGATCGGCAAGGCGCGGCTTACGTGAACCGGCGCGGCGAG 4721
Qy AGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAGCTGAGAGGCGCATTCAGCTGG 2113
Db AGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAGCTGAGAGGCGCATTCAGCTGG 4781
Qy CCTTCAGAGCAGGCGGAGAGGTTGATCTGTGACCGACAGGAGTACGCGCTTGGGCA 2173
Db CCTTCAGAGCAGGCGGAGAGGTTGATCTGTGACCGACAGGAGTACGCGCGCTTGGGCA 4841
Qy TCATTCAGGCGGAGCGGCAAGAGCGAGGAGGAGCTGTGATCAATTCAGAGAGC 2233
Db TCATTCAGGCGGAGCGGCAAGAGCGAGGAGGAGCTGTGATCAATTCAGAGAGC 4901
Qy TCATTCAGGCGGAGCGGCAAGAGCGAGGAGGAGCTGTGATCAATTCAGAGAGC 2234
Db TCATTCAGGCGGAGCGGCAAGAGCGAGGAGGAGCTGTGATCAATTCAGAGAGC 4961
Qy TGAATCAAGAGAGAGGTTACCTGAGCTGGGAGCGCGCGCAAGAGGCGATTCGCGGCA 2293
Db TGAATCAAGAGAGAGGTTACCTGAGCTGGGAGCGCGCGCAAGAGGCGATTCGCGGCA 5020
Qy AGAAGCAGATCGACAGCTGTGAGAGAGGCAATCGCAAGGAGTGTCTTCTGAGAGCA 2353
Db AGAAGCAGATCGACAGCTGTGAGAGAGGCAATCGCAAGGAGTGTCTTCTGAGAGCA 5021
Qy TGAATGCGGAGATGTGATCTTACAGATCATGAGAGCATGTTCAGTGGGCGAGCGCGCGC 2413
Db TGAATGCGGAGATGTGATCTTACAGATCATGAGAGCATGTTCAGTGGGCGAGCGCGCGC 5081
Qy CTAGAGTGAATTAAGCTTCCGCGGCGCTACACCGGT 2414

Dh 61 GCGACGACTTCAAGGAGCCCAAGGCATCATCAAGTCTTCACTGCGGCAAGAGGGC 120
Qy 121 CACATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGGCTGCTGCAAGTGCAGAGAG 180
Dh 121 CACATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGGCTGCTGCAAGTGCAGAGAG 180
Qy 181 GCGCAACGAGTGAAGGACTGCAACGAGGCGCGAGCCCACTTCTTCCGCGAGGACCTGAGC 240
Dh 181 GCGCAACGAGTGAAGGACTGCAACGAGGCGCGAGCCCACTTCTTCCGCGAGGACCTGAGC 240
Qy 241 TTTCCCGCAAGGAGGAGCGCGCGAGTTCCTCCAGCGAGCAAGCCGCGCAACGCCCAAC 300
Dh 241 TTTCCCGCAAGGAGGAGCGCGCGAGTTCCTCCAGCGAGCAAGCCGCGCAACGCCCAAC 300
Qy 301 AGCCGCGAGCTGAGAGTGCAGCGCGAGCAACCCCGCGAGGAGCGCGCGCGAGCGCGAG 360
Dh 301 AGCCGCGAGCTGAGAGTGCAGCGCGAGCAACCCCGCGAGGAGCGCGCGCGAGCGCGAG 360
Qy 361 GGCACCCCTGAACCTTCCCGCGAGTCAACCCCTGAGCGAGCGCGCGCGAGTGAAGTGA 420
Dh 361 GGCACCCCTGAACCTTCCCGCGAGTCAACCCCTGAGCGAGCGCGCGCGAGTGAAGTGA 420
Qy 421 GCGCGCGAGATCAAGAGGCGCGCTGCTGAGCAACGCGCGAGAGCAACCGTGTGAGAGAG 480
Dh 421 GCGCGCGAGATCAAGAGGCGCGCTGCTGAGCAACGCGCGAGAGCAACCGTGTGAGAGAG 480
Qy 481 ATGAGCTGCGCGCGAGAGTGAAGGCCCAAGATGATCGAGCGGATCGAGCGCTTCATCAAG 540
Dh 481 ATGAGCTGCGCGCGAGAGTGAAGGCCCAAGATGATCGAGCGGATCGAGCGCTTCATCAAG 540
Qy 541 GTCGCGCGAGTGAAGCAAGATCTGATCGAGTCTGCGCGAGCAAGAGGCCCTCGGCAACCG 600
Dh 541 GTCGCGCGAGTGAAGCAAGATCTGATCGAGTCTGCGCGAGCAAGAGGCCCTCGGCAACCG 600
Qy 601 CTGATCGCGCGAGCGCGCGTGAACATCATCGCGCGAGCAATGCTGACCGAGCTGGGCTGC 660
Dh 601 CTGATCGCGCGAGCGCGCGTGAACATCATCGCGCGAGCAATGCTGACCGAGCTGGGCTGC 660
Qy 661 ACCCTGAACCTTCCCGCATCAAGCGCGCATCGAGCCGTCGCGTGAAGCTGAAGCCCGCATG 720
Dh 661 ACCCTGAACCTTCCCGCATCAAGCGCGCATCGAGCCGTCGCGTGAAGCTGAAGCCCGCATG 720
Qy 721 GACGCGCGCGAGTGAAGAGTGGCGCGCTGACCGAGAGAGAAATCAAGGCCCTGACCGCG 780
Dh 721 GACGCGCGCGAGTGAAGAGTGGCGCGCTGACCGAGAGAGAAATCAAGGCCCTGACCGCG 780
Qy 781 ATCTGAGAGAGTGAAGAGGAGGAGCAAGTCAACAGTGGCGCGCGAGAACCGCTTAC 840
Dh 781 ATCTGAGAGAGTGAAGAGGAGGAGCAAGTCAACAGTGGCGCGCGAGAACCGCTTAC 840
Qy 841 AACACCCCGCTGTTCGCAATCAAGAGAGAGCAAGCAACAGTGGCGCGAGCTGTGAGAG 900
Dh 841 AACACCCCGCTGTTCGCAATCAAGAGAGAGCAAGCAACAGTGGCGCGAGCTGTGAGAG 900
Qy 901 TTTCCGCGAGCTGAACAGCGCAACCGAGACTTCTGAGAGTGCAGCTGGGCAATCCCCAC 960
Dh 901 TTTCCGCGAGCTGAACAGCGCAACCGAGACTTCTGAGAGTGCAGCTGGGCAATCCCCAC 960
Qy 961 CCGCGCGGCTGAAGAGAGAGAGAGAGAGTGAACCGTGTGAGAGTGGCGAGCGCTTACTTC 1020
Dh 961 CCGCGCGGCTGAAGAGAGAGAGAGAGTGAACCGTGTGAGAGTGGCGAGCGCTTACTTC 1020
Qy 1021 AGCGTGCCTGTGAGAGAGACTTTCGCAAGTACACCGCTTCACATCCCGAGCATCAAC 1080
Dh 1021 AGCGTGCCTGTGAGAGAGACTTTCGCAAGTACACCGCTTCACATCCCGAGCATCAAC 1080
Qy 1081 AAGAGAGAGAGAGAGAGAGAGAGAGAGTGAAGAGTGTGCGAGGCTGAGAGAGAGAGAG 1140
Dh 1081 AAGAGAGAGAGAGAGAGAGAGAGAGAGTGAAGAGTGTGCGAGGCTGAGAGAGAGAGAG 1140
Qy 1141 CCGAGCATCTTCAAGAGAGAGATGAACAGATCTTGAAGCCCTTCGCGCGCGAGAACCC 1200
Dh 1141 CCGAGCATCTTCAAGAGAGAGATGAACAGATCTTGAAGCCCTTCGCGCGCGAGAACCC 1200

Qy 1201 GAGATGATGATTAACAGGCGCGCGCTGTACGTTGGGCAAGGACCTGAGATGGGCGAGC 1260
Dh 1201 GAGATGATGATTAACAGGCGCGCGCTGTACGTTGGGCAAGGACCTGAGATGGGCGAGC 1260
Qy 1261 GCGCGCAAGATCAAGAGAGTGGCGCAAGCACTGTGCGTGGGCTTCACACCCCGAGC 1320
Dh 1261 GCGCGCAAGATCAAGAGAGTGGCGCAAGCACTGTGCGTGGGCTTCACACCCCGAGC 1320
Qy 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
Dh 1321 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
Qy 1375 TGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
Dh 1375 TGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
Qy 1435 AAGCTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1494
Dh 1441 AAGCTGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy 1495 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGTGCGCTGACCGAGAG 1554
Dh 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGTGCGCTGACCGAGAG 1560
Qy 1555 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614
Dh 1561 GCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1615 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
Dh 1621 TACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1675 TACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
Dh 1681 TACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy 1735 ACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Dh 1741 ACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1795 AGCATGTGATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Dh 1801 AGCATGTGATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1855 GAGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Dh 1861 GAGACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy 1915 ACCCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
Dh 1921 ACCCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1975 ACCCTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
Dh 1981 ACCCTTCAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy 2035 ACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
Dh 2041 ACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy 2095 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
Dh 2101 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy 2155 AGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
Dh 2161 AGCGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2215 AACGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
Dh 2221 AACGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280

QY 2275 CACAAGGCGCATGCGCGGCAACGAGCATGCAAGCTGTGTAGCAAGGCGATCCGCAAG 2334
DB 2281 CACAAGGCGCATGCGCGGCAACGAGCATGCAAGCTGTGTAGCAAGGCGATCCGCAAG 2340
QY 2335 GTGCTGTTCTGTGACGCGCATGCGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATG 2394
DB 2341 GTGCTGTTCTGTGACGCGCATGCGCGCATGCGCATGCGCATGCGCATGCGCATGCGCATG 2400
QY 2395 TACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2451
DB 2401 TACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2457

RESULT 13
ADCL3265
ID ADCL3265 standard; DNA; 2457 BP.
XX ADCL3265;
AC
XX
XX 18-DEC-2003 (first entry)
DT
XX
DE DNA of HIV construct p2Pol-opt-YM_C SBQ ID NO 44.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Pro; Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
XX Human immunodeficiency virus.
XX
XX WO2003004620-A2.
XX
XX 16-JAN-2003.
PD
XX
XX 05-JUL-2002; 2002WO-US021420.
PF
XX
XX 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
XX
XX (CHIR) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megede J, Barnette SM, Llan Y, Engelbrecht S, Van Rensburg EJ;
XX
XX MPI, 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Pro, or Rev polypeptide, useful for immunization, or generating packaging cell lines.
XX
XX
XX Disclosure; Fig 41; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a polynucleotide sequence encoding a polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat, Pro, or Rev polypeptide. The novel expression cassette can be used to treat HIV type C by gene therapy or used in the development of a vaccine. The gene delivery vector is administered intramuscularly, intramuscularly, intravenously, subcutaneously, intradermally, transdermally, intranasally, intrarectally, orally or intravenously. The expression cassette is useful for immunisation, generating packaging cell lines and producing HIV polypeptides. This polynucleotide sequence represents the DNA of an HIV Type C related sequence of the invention.
XX
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.8%; Score 2428.6; DB 10; Length 2457;
Best Local Similarity 99.6%; Pred No. 3.6e-292;
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 1 GTGACGCGCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60
DB 1 GTGACGCGCCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60

QY 61 CGCAGCACTTCAAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
DB 61 CGCAGCACTTCAAGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 121 CACATCGCCCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 121 CACATCGCCCGCAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 GCGCACCAGATGAAGAGCTGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
DB 181 GCGCACCAGATGAAGAGCTGCAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
QY 241 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
DB 241 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
QY 301 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 301 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 GCGACCTGGAATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
DB 361 GCGACCTGGAATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 421 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
DB 421 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 481 ATGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
DB 481 ATGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
DB 541 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 601 CTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGGAATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 ACCCTGGAATTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 781 ATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
DB 781 ATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 841 AACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
DB 841 AACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
QY 901 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
DB 901 TTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
QY 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
QY 1021 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 1021 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 1081 AACGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 1081 AACGAGACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140

QY 1141 CCCAGCATCTTCAGAGCAGCATGAACAGATCCTGAGCCCTTCGGGCGCCGCAACCC 1200
DB 1141 CCCAGCATCTTCAGAGCAGCATGAACAGATCCTGAGCCCTTCGGGCGCCGCAACCC 1200
QY 1201 GAGATCGTGAATCTACCAAGGCCCCCTGTACTGTGGCAGGACCTGGAAGATCGGGCAGCAG 1260
DB 1201 GAGATCGTGAATCTACCAAGGCCCCCTGTACTGTGGCAGGACCTGGAAGATCGGGCAGCAG 1260
QY 1261 CGCGCCAGAGATGAGAGAGCTGGCAGAGCACTGTGCGGTGGGCTTCCACACCCCGCAG 1320
DB 1261 CGCGCCAGAGATGAGAGAGCTGGCAGAGCACTGTGCGGTGGGCTTCCACACCCCGCAG 1320
QY 1321 AAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1374
DB 1321 AAGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1375 TGGACCGTGCAGACCCCATCGAGCTGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
DB 1381 TGGACCGTGCAGACCCCATCGAGCTGCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1435 AAGCTGTGGGCAAGCTGAATCTGAGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCCAG 1494
DB 1441 AAGCTGTGGGCAAGCTGAATCTGAGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCCAG 1500
QY 1495 CTGTGCAAGCTGTGTGGCGCGCCGCAAGGCGCTGACCCGACATGTGTGCGCTTGAACGAGAG 1554
DB 1501 CTGTGCAAGCTGTGTGGCGCGCCGCAAGGCGCTGACCCGACATGTGTGCGCTTGAACGAGAG 1560
QY 1555 GCCGAGCTGAGAGCTGGCCGAGAGACCGAGAGATCTGTGCGGAGCCCGTGACGAGCGGTGAC 1614
DB 1561 GCCGAGCTGAGAGCTGGCCGAGAGACCGAGAGATCTGTGCGGAGCCCGTGACGAGCGGTGAC 1620
QY 1615 TACGACCCCGAGAGAGAGCTGTGTGCGCAGATCTCAGAGCAGGCGCCAGCAGAGTGGAGC 1674
DB 1621 TACGACCCCGAGAGAGAGCTGTGTGCGCAGATCTCAGAGCAGGCGCCAGCAGAGTGGAGC 1680
QY 1675 TACGAGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
DB 1681 TACGAGATCTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1735 ACCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
DB 1741 ACCGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1795 AAGCATGTGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
DB 1801 AAGCATGTGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1855 GAGACCTGTGAGACCGACTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
DB 1861 GAGACCTGTGAGACCGACTACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1915 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
DB 1921 ACCCCCCCTGTGTGAAGCTGTGTGTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1975 ACCCTTCAAGTGAACGCGCCCGCAACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
DB 1981 ACCCTTCAAGTGAACGCGCCCGCAACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2035 ACCGACCGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
DB 2041 ACCGACCGGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2095 CTGCAAGGAGATCGAGCTGCGCTGCGAGCAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
DB 2101 CTGCAAGGAGATCGAGCTGCGCTGCGAGCAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2155 AGCCAGTAAGCCCTGTGGGAGATCATTCAGGCGCCAGCCGAGCAAGAGAGAGAGAGAGAGAGAGAG 2214
DB 2161 AGCCAGTAAGCCCTGTGGGAGATCATTCAGGCGCCAGCCGAGCAAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2215 AACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274

DB 2221 AACCAAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2275 CACAAAGGAGATCGGCGGCAAGAGAGATCGAACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334
DB 2281 CACAAAGGAGATCGGCGGCAAGAGAGATCGAACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
QY 2335 GTGCTGTCTTGGAGAGGAGATGATGAGGCGGAGATCTGATCTACAGATACATGAGAGAGAGAGAGAG 2394
DB 2341 GTGCTGTCTTGGAGAGGAGATGATGAGGCGGAGATCTGATCTACAGATACATGAGAGAGAGAGAGAG 2400
QY 2395 TACGTTGGCAGCGGCGGCGCTAGAGATCGATTAAAGCTTCCGCGGCTAGCAGCCGCT 2451
DB 2401 TACGTTGGCAGCGGCGGCGCTAGAGATCGATTAAAGCTTCCGCGGCTAGCAGCCGCT 2457
RESULT 14
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnette SM, Engelbrecht S, Van Rensburg EJ;
XX
DR MPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (i) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;
Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.5e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
QY 1 GTGACGCGCAGCATGAGCGAGGAGCATGAGCCAGGCGCAGCAGGCGCCAGCATCTGTGAGAG 60

Db 1 GTGACGCGACCATGAGCGAGGCGCATGAGCCAGGCGCAACAGCGCCAAATCTGTATGAG 60
Qy 61 CGGAGCAATCTTCAAGAGGCGCCCAAGGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCG 120
Db 61 CGGAGCAATCTTCAAGAGGCGCCCAAGGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCG 120
Qy 121 CACATGCGCGGCACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGTGCGCAAGAG 180
Db 121 CACATGCGCGGCACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGTGCGCAAGAG 180
Qy 181 GGCACACGATGAAAGATGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGCGCC 240
Db 181 GGCACACGATGAAAGATGCAACCGAGCGCCAGGCGCAACTTCTTCGCGAGGACCTGCGCC 240
Qy 241 TTCCCGCGAGGCGCAAGGCGCGAGTTCGCCAGAGGAGAAACCGAGCGCAACAGCGCCGAC 300
Db 241 TTCCCGCGAGGCGCAAGGCGCGAGTTCGCCAGAGGAGAAACCGAGCGCAACAGCGCCGAC 300
Qy 301 AGCCGCGAGCTGCAAGTGTGCGGCGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCG 360
Db 301 AGCCGCGAGCTGCAAGTGTGCGGCGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCGCG 360
Qy 361 GGCACCTGTAACTTCCCCAGATCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 420
Db 361 GGCACCTGTAACTTCCCCAGATCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCG 420
Qy 421 GCGCGCGAGATCAAGAGGCGCGTGTGCAACCGCGCGCGAGCAACCGTGTGTGAGAG 480
Db 421 GCGCGCGAGATCAAGAGGCGCGTGTGCAACCGCGCGCGAGCAACCGTGTGTGAGAG 480
Qy 481 ATGAGCTGTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAG 540
Db 481 ATGAGCTGTGCGCGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAACAG 540
Qy 541 GTGCGCGAGTACCAAGATCTGTATCGAGATCTGTGTGCGAGAGAGCGCATCGCGCAC 600
Db 541 GTGCGCGAGTACCAAGATCTGTATCGAGATCTGTGTGCGAGAGAGCGCATCGCGCAC 600
Qy 601 CTGATGTGCGCGCGCGCGCGTGAACATCATGTGCGCGCGCAACATCGTGTGTGCGCG 660
Db 601 CTGATGTGCGCGCGCGCGCGCGTGAACATCATGTGCGCGCGCAACATCGTGTGTGCGCG 660
Qy 661 ACCCTGAATCTTCCCATCAGCGCCCATTCAGAACCGTGTGCGCGCGCGCGCGCGCG 720
Db 661 ACCCTGAATCTTCCCATCAGCGCCCATTCAGAACCGTGTGCGCGCGCGCGCGCGCG 720
Qy 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Qy 781 ATCTGCGAGAGATGGAAGAGGCGCAAGATCAACAGATGTGCGCGCGCGCGCGCGCG 840
Db 781 ATCTGCGAGAGATGGAAGAGGCGCAAGATCAACAGATGTGCGCGCGCGCGCGCGCG 840
Qy 841 AACACCGCGGTGTGCGCATCAAGAGAGAGCAGACCAAGTGTGCGCGCGCGCGCGCG 900
Db 841 AACACCGCGGTGTGCGCATCAAGAGAGAGCAGACCAAGTGTGCGCGCGCGCGCGCG 900
Qy 901 TTCCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGTGAGGTGCACTGTGCGCGCAC 960
Db 901 TTCCCGCGAGCTGAAACAGCGCACCCAGGACTTCTGTGAGGTGCACTGTGCGCGCAC 960
Qy 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Db 961 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
Qy 1021 AGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Db 1021 AGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
Qy 1081 AACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Db 1081 AACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140

Db 1081 AACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
Qy 1141 CCGAGCATCTTCAAGAGGCGATGACCAAGATCTGTGAGCGCGCTTCCGCGCGCGCAAC 1200
Db 1141 CCGAGCATCTTCAAGAGGCGATGACCAAGATCTGTGAGCGCGCTTCCGCGCGCGCAAC 1200
Qy 1201 GAGATGTGTATCTTCA-----GCGCGCGGTGATCGTGTGCGAGCGACCTGTGAGTGTG 1254
Db 1201 GAGATGTGTATCTTCAAGATGATGAGACCGATCGTGTGCGAGCGACCTGTGAGTGTG 1260
Qy 1255 CAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1314
Db 1255 CAGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
Qy 1315 CCGGCAAGAGAGCACAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1368
Db 1315 CCGGCAAGAGAGCACAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
Qy 1369 GACAAGTGAGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1428
Db 1369 GACAAGTGAGCGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1440
Qy 1429 ATCCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1488
Db 1441 ATCCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
Qy 1489 GCGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1548
Db 1501 GCGCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
Qy 1549 GAGAGGCGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1608
Db 1561 GAGAGGCGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
Qy 1609 GTGTACTTACGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1668
Db 1621 GTGTACTTACGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1669 TGAGACTTACGAGATCTTACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1728
Db 1681 TGAGACTTACGAGATCTTACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1740
Qy 1729 ATGCGCACCGCGCGCACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1788
Db 1741 ATGCGCACCGCGCGCACCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy 1789 ATGAGAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1848
Db 1801 ATGAGAGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Qy 1849 ACTGTGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1908
Db 1861 ACTGTGAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
Qy 1909 GTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1968
Db 1921 GTGAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
Qy 1969 GCGGAGACCTTCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2028
Db 1981 GCGGAGACCTTCTTACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
Qy 2029 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2088
Db 2041 TACGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2100
Qy 2089 ACCGAGCTGCAAGGCGCATCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2148
Db 2101 ACCGAGCTGCAAGGCGCATCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2160
Qy 2149 ACCGAGCGCGAGTACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2208
Db 2161 ACCGAGCGCGAGTACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220

QY 2209 CTGTGAACGAGATCATGAGCACTGATCAAGAGAGAGTGTACTGAGCTGGTG 2268
 DB 2221 CTGTGAACGAGATCATGAGCACTGATCAAGAGAGAGTGTACTGAGCTGGTG 2280
 QY 2259 CCGGCGCAAGAGGAGTGGCGGCAACGAGATCGAAGCTGTGAGCAAGGGCATC 2328
 DB 2281 CCGGCGCAAGAGGAGTGGCGGCAACGAGATCGAAGCTGTGAGCAAGGGCATC 2340
 QY 2329 CGCAAGTGTCTGTCTGTGACGCGATCGATGCGGCGATCTGATCAACGATGACATG 2388
 DB 2341 CGCAAGTGTCTGTCTGTGACGCGATCGATGCGGCGATCTGATCAACGATGACATG 2400
 QY 2389 GACCTGTACGTGGGCAAGCGCGCGCTTAGATCGATTAAAGCTTCCGGGCTAGCAC 2448
 DB 2401 GACCTGTACGTGGGCAAGCGCGCGCTTAGATCGATTAAAGCTTCCGGGCTAGCAC 2460
 QY 2449 GGTGAATTC 2457
 DB 2461 GGTGAATTC 2469

RESULT 15
 ADM73764
 ID ADM73764 standard; DNA; 2469 BP.
 AC ADM73764;
 DT 03-JUN-2004 (first entry)
 DE HIV-1 polynucleotide #7.
 XX HIV-1 polynucleotide #7.
 KM HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
 XX HIV type C protein; immunostimulant.
 OS Human immunodeficiency virus 1.
 PN US2003223961-A1.
 PD 04-DEC-2003.
 PF 05-JUL-2001; 2001US-00899575.
 PR 05-JUL-2000; 2000US-00610313.
 PA (MEGE/) MEGEDE J Z.
 PA (BARRE/) BARRETT S W.
 PA (ENG/) ENGELBRECHT S.
 PA (RENS/) RENSBURG R J V.
 PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg RJV;
 XX WPI; 2004-060515/06.
 DR
 XX
 PT New expression cassette comprising a polynucleotide sequence encoding an
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
 PT immunization, generating of packaging cell lines or in producing HIV Type
 PT C proteins.
 PS Claim 1; SEQ ID NO 30; 160bp; English.
 XX
 CC The invention relates to an expression cassette comprising a
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
 CC also relates to a recombinant expression system for use in a host cell
 CC comprising an expression cassette, where the polynucleotide sequence
 CC further comprises control elements capable of driving expression in the
 CC selected host cell, a cell comprising an expression cassette where the
 CC polynucleotide sequence further comprises control elements compatible
 CC with the expression in the cell and a composition for generating an
 CC immunological response, comprising an expression cassette. The expression
 CC cassette and the methods of the invention are useful in eliciting an
 CC immune response, in DNA immunisation, in generation of packaging cell
 CC lines and in producing HIV Type C proteins. This sequence represents an

CC HIV-1 polynucleotide of the invention.
 XX SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;
 Query Match 98.3%; Score 2415.4; DB 12; Length 2469;
 Best Local Similarity 99.3%; Pred. No. 1.5e-290;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 1 GTGACGCGCACATGCGCGGAGCCATGAGCCGACAGGCGCACAGCCGACATCTCTGATGAG 60
 DB 1 GTGACGCGCACATGCGCGGAGCCATGAGCCGACAGGCGCACAGCCGACATCTCTGATGAG 60
 QY 61 CGGAGCAACTTCAAGGGGCGGCAAGGAGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120
 DB 61 CGGAGCAACTTCAAGGGGCGGCAAGGAGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120
 QY 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAACTGCGGCAAGAG 180
 DB 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAACTGCGGCAAGAG 180
 QY 181 GCGCACACGATGAGAGGATCTGACCGAGCGCCAGGCGCACTTCTTCCGAGAGACTGTGGCC 240
 DB 181 GCGCACACGATGAGAGGATCTGACCGAGCGCCAGGCGCACTTCTTCCGAGAGACTGTGGCC 240
 QY 241 TTCCCGGAGGCGCAAGGCGCGGAGTTCCCGAGGAGCAACCGGCGCAAGCGCCAC 300
 DB 241 TTCCCGGAGGCGCAAGGCGCGGAGTTCCCGAGGAGCAACCGGCGCAAGCGCCAC 300
 QY 301 AGCCGAGCTGACAGGTGCGCGGCGACCAACCGCGCGACAGGCGCGCGCGCGCGCG 360
 DB 301 AGCCGAGCTGACAGGTGCGCGGCGACCAACCGCGCGACAGGCGCGCGCGCGCGCGCG 360
 QY 361 GGCACCTGAACTTCCCGCAAGTCACTCTGTGAGCGCGCGCTGTGTAGCATCAAGTG 420
 DB 361 GGCACCTGAACTTCCCGCAAGTCACTCTGTGAGCGCGCGCTGTGTAGCATCAAGTG 420
 QY 421 GCGGCGCAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGTGTGAGAG 480
 DB 421 GCGGCGCAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGTGTGAGAG 480
 QY 481 ATGAGCTGCGCGGAGAGTGAAGCCCAAGATGATCGGCGGATCGGCGCTTCAATCAAG 540
 DB 481 ATGAGCTGCGCGGAGAGTGAAGCCCAAGATGATCGGCGGATCGGCGCTTCAATCAAG 540
 QY 541 GTGCGCGAGTACAGCAAGATCTGTATGAGATCTGCGGCAAGAGGCGCATCGGACCGTG 600
 DB 541 GTGCGCGAGTACAGCAAGATCTGTATGAGATCTGCGGCAAGAGGCGCATCGGACCGTG 600
 QY 601 CTGATCGGCGCGCAACCGCGGTGAACATCATCGGCGCGCAATGCTGAACCCAGCTGGGCTGC 660
 DB 601 CTGATCGGCGCGCAACCGCGGTGAACATCATCGGCGCGCAATGCTGAACCCAGCTGGGCTGC 660
 QY 661 ACCCTGAATCTCCCATGAGCGCCCATCGAGACCGTGCCTGTAGCTGAAGCCCGGATG 720
 DB 661 ACCCTGAATCTCCCATGAGCGCCCATCGAGACCGTGCCTGTAGCTGAAGCCCGGATG 720
 QY 721 GACGCGCGCGCAAGTGAAGAGTGGCGCTGTGACCGAGAGAGATCAAGGCGCTGTGACG 780
 DB 721 GACGCGCGCGCAAGTGAAGAGTGGCGCTGTGACCGAGAGAGATCAAGGCGCTGTGACG 780
 QY 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCGAGAACCCCTTAC 840
 DB 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCGAGAACCCCTTAC 840
 QY 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 DB 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 901 TTCGCGAGCTGAACAGGCGACCCAGAGACTTCTGTGAGGTGACGTGGGCAATCCCGCAC 960
 DB 901 TTCGCGAGCTGAACAGGCGACCCAGAGACTTCTGTGAGGTGACGTGGGCAATCCCGCAC 960
 QY 961 CCGCGGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020

```

Db 961 CCGCGCGGCTGTGAAGAGAGAGCGTGAACGTGTGAGTGTGGGAGACGCTACTTCC 1020
Qy 1021 AGCGTGGCCCTGTGAAGAGAGCTTCCGCAAGTACACCGGCTTTCACATCCCGGACATCAAC 1080
Db 1021 AGCGTGGCCCTGTGAAGAGAGCTTCCGCAAGTACACCGGCTTTCACATCCCGGACATCAAC 1080
Qy 1081 AACGAGACCCCGGAGATCCGCTACCAAGTACCAAGTGTGCCCCGAGGCTGTGAAGGAGAC 1140
Db 1081 AACGAGACCCCGGAGATCCGCTACCAAGTACCAAGTGTGCCCCGAGGCTGTGAAGGAGAC 1140
Qy 1141 CCGAGCATCTTTCAGAGCAGCAGTGAACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCC 1200
Db 1141 CCGAGCATCTTTCAGAGCAGCAGTGAACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCC 1200
Qy 1201 GAGATCGTGAATCTACA-----GGCCCCCTGTGAGTGTGGGAGAGACCTGTGAATCGGC 1254
Db 1201 GAGATCGTGAATCTACAAGTACATGTGACGACTGTGAGTGTGGGAGAGACCTGTGAATCGGC 1260
Qy 1255 CAGCACCGGCGCAAGATCGAGAGAGCTGTGCGCAAGACCTGTGCGCTGGGGCTTTCACACC 1314
Db 1261 CAGCACCGGCGCAAGATCGAGAGAGCTGTGCGCAAGACCTGTGCGCTGGGGCTTTCACACC 1320
Qy 1315 CCGGACAAAGAGACCAAGAGAGCCCTTCTGTGCGCAT-----CGAGCTGACCCC 1368
Db 1321 CCGGACAAAGAGACCAAGAGAGCCCTTCTGTGAGTGTGGCTACGAGCTGACCCC 1380
Qy 1369 GACAGTGTGACCGTGTGAGCGCCATCGAGCTGTGCGGAGAGAGAGAGCTGTGACCGTGAACAC 1428
Db 1381 GACAGTGTGACCGTGTGAGCGCCATCGAGCTGTGCGGAGAGAGAGAGCTGTGACCGTGAACAC 1440
Qy 1429 ATTCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1488
Db 1441 ATTCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
Qy 1489 CGCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1548
Db 1501 CGCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560
Qy 1549 GAGAGAGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1608
Db 1561 GAGAGAGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1620
Qy 1609 GTGTACTAGACCCCGAGCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1668
Db 1621 GTGTACTAGACCCCGAGCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1680
Qy 1669 TGGACCTACAGATCTACAGAGAGCCCTTCAAGAACTGTGAAGCCGGCAAGTACGCCAAG 1728
Db 1681 TGGACCTACAGATCTACAGAGAGCCCTTCAAGAACTGTGAAGCCGGCAAGTACGCCAAG 1740
Qy 1729 ATGTGCGACCGGCGCAACCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1788
Db 1741 ATGTGCGACCGGCGCAACCAAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800
Qy 1789 ATGTGAGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1848
Db 1801 ATGTGAGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860
Qy 1849 ACCTGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1908
Db 1861 ACCTGTGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920
Qy 1909 GTGAACAACCCCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1968
Db 1921 GTGAACAACCCCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Qy 1969 GCGGAGACCTTCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2028
Db 1981 GCGGAGACCTTCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
Qy 2029 TACGTGACCGAGCCGGGCGGAGAGATGTGTGACCTGTGACCGAGACCAACCAAGAG 2088

```

```

Db 2041 TACGTGACCGAGCCGGGCGGAGAGATGTGTGACCTGTGACCGAGACCAACCAAGAG 2100
Qy 2089 ACCGAGCTGTGAGGCGCATTCAGCTGTGCGCTGTGAGGACAGCGGCGAGGAGTGAACATGTGT 2148
Db 2101 ACCGAGCTGTGAGGCGCATTCAGCTGTGCGCTGTGAGGACAGCGGCGAGGAGTGAACATGTGT 2160
Qy 2149 ACCGACAGCAGTACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2208
Db 2161 ACCGACAGCAGTACGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
Qy 2209 CTGTGTAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2268
Db 2221 CTGTGTAACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2280
Qy 2269 CCGGCGCAAGAGGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2328
Db 2281 CCGGCGCAAGAGGCGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2340
Qy 2329 CGCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2388
Db 2341 CGCAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2400
Qy 2389 GACCTGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2448
Db 2401 GACCTGTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2460
Qy 2449 GTGTGAATTC 2457
Db 2461 GTGTGAATTC 2469

```

Search completed: December 30, 2005, 08:57:08
 Job time : 1300.51 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8676.48 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457
1 gtgcagccacacatgcccga.....gggctagcaccggtgaatc 2457

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*
10: gb_esc10:*
11: gb_esc11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136.4	5.6	330	10	CL293849 02S0349-0
2	100	4.1	2886	10	CL967755 OIRFC0015
3	90.2	3.7	1781	8	CKX096623 RECM0260
4	90.2	3.7	2598	4	AY103647 Zea mayb
5	88.8	3.6	3069	10	CL973391 OIRFC0025
6	86.6	3.5	1398	10	CL961989 OIRFC0006
7	86	3.5	743	10	CZ247380 AIAA-aaf3
8	85.6	3.5	951	3	BM321451 rockefell
9	84.2	3.4	1941	10	CL971508 OIRFC0021
10	80.8	3.3	1060	10	CW922203 EDGAR29TR
11	80.6	3.3	869	7	CK159167 FAS04056
12	80.2	3.3	892	10	CZ216254 AIAA-aaf2
13	80.2	3.3	1132	3	BM320864 rockefell
14	79.6	3.2	1165	3	BM320900 rockefell
15	79.6	3.2	1680	10	CL982270 OIRFC0049
16	79.6	3.2	867	3	BM321430 rockefell
17	78.4	3.2	1725	10	CL978463 OIRFC0031
18	78.4	3.2	2031	10	CL974989 OIRFC0042
19	78.2	3.2	2697	10	CL952258 OIRFC0000
20	78.2	3.2	2853	10	CL974397 OIRFC0025
21	78	3.2	1485	10	CL970981 OIRFC0020
22	77.4	3.2	1509	10	CL959255 OIRFC0002

23	77.4	3.2	2559	10	CL982027 OIRFC0046
24	77	3.1	11691	10	CL962901 OIRFC0008
25	76.8	3.1	2682	10	CL966033 OIRFC0017
26	76.6	3.1	1550	3	BM321022 rockefell
27	76.2	3.1	1401	10	CL962721 OIRFC0038
28	76.2	3.1	3249	10	CL945510 OIRFC0004
29	75	3.1	545	3	BM321023 rockefell
30	75	3.1	862	3	BM321023 rockefell
31	74.4	3.0	914	9	BZ568300 pac82-164
32	74.2	3.0	2313	10	CL982362 OIRFC0047
33	74	3.0	1290	10	CL972679 OIRFC0023
34	74	3.0	1386	11	DQ045165 Homo sapi
35	74	3.0	2072	4	CR603312 full-1eng
36	73.8	3.0	2433	10	AY401196 Homo sapi
37	73.2	3.0	853	3	BM321393 rockefell
38	72.8	3.0	566	3	BM587428
39	72.6	3.0	718	4	CNS06GHPN
40	72.6	3.0	788	6	CB643171 OIRFC003L
41	72.4	2.9	753	9	CC675888 OIRFC001TH
42	72.4	2.9	2523	10	CL974879 OIRFC0026
43	72.2	2.9	602	7	CV057146 BNE124e3
44	72.2	2.9	640	2	BE601575 HVSMBH009
45	72.2	2.9	688	6	CB648640 OSJNB12C

ALIGNMENTS

RESULT 1
CL293849 330 bp DNA linear GSS 12-FEB-2004
LOCUS 02S0349-08A1-C03 UniformMu MutRIL Library Zea mays genomic clone
DEFINITION 02S0349-08A1-C03, genomic survey sequence.
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 330)
Lathaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
Sequence tagged transposon insertions from the UniformMu maize population
Unpublished (2003)
JOURNAL Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmcufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
02S0349-08, Primer set: A
Classes: transposon insertion site.
Location/Qualifiers
1..330

FEATURES
source
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric PCR (TAP) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sephadex 400 spin columns and cloned into the TOPO PCR4 vector."

ORIGIN

Query Match 5.6%; Score 136.4; DB 10; Length 330;
Best Local Similarity 66.4%; Pred. No. 3.1e-14;
Matches 211; Conservative 0; Mismatches 106; Indels 1; Gaps 1;

QY 421 GGGGGCCGATTCAGAGAGCCCTGCTGACACCGGCGCCGACGACCGTGTGAGAG 480
DB |||||
QY 339 GGGGGGCGAGCTGAGAGAGCTCTATTATGATACAGAGCAGATGATACGATATTAAAGAA 270
DB |||||
QY 481 ATGAGCCTGCGCGGAGAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCAAG 540
DB |||||
QY 269 ATGACTTTGACAGAGAGATGAGAACCAAAATGATAGGGGAAATTGAGGTTTATCAAA 210
DB |||||
QY 541 GTGCGCCGATGACGACCAAGATCTGATTCGAGATCTCGGCGAGAGAGCCATCGGACCGTG 600
DB |||||
QY 209 GTAAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 150
DB |||||
QY 601 CTGATCGGCGCCGACCGCGTGAACATCATTCGCGCCGACATCTGATCCGATGAGGCTGC 660
DB |||||
QY 149 TTAGTAGAGACTTACACCTGTCTCAACATTAATGAGAAATCTGTGATCTCAATTTGG-TGC 91
DB |||||
QY 661 ACCCTGAACCTTCCCATGACCCCATGACGCGTGCCTGAGAGCTGAAAGCCCGGCAATG 720
DB |||||
QY 90 ACCTTAAATTTTCCCATTTAGTCTTATTTGAACTGATACGATTAATTAAGCCAGGAATG 31
DB |||||
QY 721 GACGGCCCGCAAGGTGAAG 738
DB |||||
QY 30 GATGGCCCAAAAGTAAG 13
DB |||||

RESULT 2
LOCUS CL967755 2886 bp DNA linear GSS 21-SEP-2004
DEFINITION OslrCCO15718 Oryza sativa Expressed Library Oryza sativa (indica
cultivar-group) genomic, genomic survey sequence.

ACCESSION CL967755.1 GI:52390149
VERSION GSS.

SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
KEYWORDS Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Rhizarioideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 2886)
AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)

TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Chinese Academy of Genomics
Beijing Institute of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

FEATURES
source
1..2886
Location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_1ib="Oryza sativa Expressed Library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 4.1%; Score 100; DB 10; Length 2886;
Best Local Similarity 42.8%; Pred. No. 1.2e-07;
Matches 700; Conservative 0; Mismatches 915; Indels 21; Gaps 3;

QY 189 GATGAGAGCTGACCGGCGCGGCGCAACTTCTTCCGCGAGAGCTTGCTTCCCA 248
DB |||||
QY |||||
DB |||||

DB 219 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 278
QY 249 GGGGAGAGCGCGCGAGTTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 308
DB |||||
QY 219 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 338
DB |||||
QY 309 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 365
DB |||||
QY 339 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 398
DB |||||
QY 366 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 425
DB |||||
QY 399 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 458
DB |||||
QY 426 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 485
DB |||||
QY 459 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 518
DB |||||
QY 486 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 545
DB |||||
QY 519 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 578
DB |||||
QY 546 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 605
DB |||||
QY 579 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 638
DB |||||
QY 606 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 665
DB |||||
QY 639 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 698
DB |||||
QY 666 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 725
DB |||||
QY 699 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 758
DB |||||
QY 726 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 785
DB |||||
QY 759 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 818
DB |||||
QY 786 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 845
DB |||||
QY 819 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 878
DB |||||
QY 846 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 905
DB |||||
QY 879 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 938
DB |||||
QY 906 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 965
DB |||||
QY 939 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 998
DB |||||
QY 966 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1025
DB |||||
QY 999 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1058
DB |||||
QY 1026 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1085
DB |||||
QY 1059 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1118
DB |||||
QY 1086 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1145
DB |||||
QY 1119 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1178
DB |||||
QY 1146 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1205
DB |||||
QY 1179 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1226
DB |||||
QY 1206 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1265
DB |||||
QY 1227 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1286
DB |||||
QY 1266 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1325
DB |||||
QY 1287 GGTGATGAGAGCGCGCGCATATGCGCATGCTCTCGCAACGCGCGGCGGCGCGCC 1346
DB |||||

Oy	1326	GCACCAAGAGAGGCC	CCCCCTTCCTGCCCATGCACTGCACCCCGACAAGATGGAACCGTGC	1385
Oy	1347	CGACCAATGACCGGA	CCCGCGCTCCCGGTGCTGAGCGGTGTGTATCCAGAAAGTGC	1406
Oy	1386	GCCCATCAGCTGC	CCCGAGAGAGAGATGTGACCGTGAACGATCCAGAAAGCTGTGTGG	1445
Oy	1407	GAAAGCAAGAGAGAC	CGAGGGCGAGCCATGGCACTTCGTGGGCTCTCTGCGCTGTTCGA	1466
Oy	1446	CAAGCTGAATGTGGG	CAGCCAGATCTACCCCGCATCAAGGTGGCCAGCTGTGCAAGCT	1505
Oy	1467	CCCGCCAGAGGACA	CGACAGCGGAGAGACCATCCGCGCGGTGTGACCTCTCGCGTGAAGCT	1526
Oy	1506	GCTGCGCGCGCC	CAAGGCTTCGACCGACATGTGTGCTCCCTGACCGAAGAGCGGAGTGG	1565
Oy	1527	GAAATGATGTCA	CCGGCGACAGCTCGCATCG-----GGAAGGAGACGGCGCGCGGCT	1580
Oy	1566	GCTGGCGAGAA	CCGCGAGATCTGTGCGGAGACCCGTGACGGCGGTGTATCTACGACCCAG	1625
Oy	1581	CGGATGTGGGACA	CATGTACCGGTGACAGACGTGTGTGGCGACAAGACAGACAGAT	1640
Oy	1626	CAAGGACCTGTGGG	CGAGATCCAGAACAGGCGCACAGACAGTGGACCTTACAGATCTA	1685
Oy	1641	GAGCGGCTCCCAAT	CGAGACTGATGAGAGAGCGGACCGCTTCGCGCGCTGTTC	1700
Oy	1686	CCAGGAGCCCTTCA	AGAACTGAAGACCGGCAAGTACGCCAAGATGTGCACCGCGCAC	1745
Oy	1701	GGAAGCAAGTAC	GAGATCTGTGAAGCGCTCAAGAACCGAAGACCAATCTGTGGGATGAC	1760
Oy	1746	CAACGACGTGAAG	CACTTGACCGAGGCGGTGCAGAAATTCGCATGAGAGCATCTGAT	1805
Oy	1761	CGGACGCGCGTGA	ACGACGCGCGCTGAAGAGGCGCATTCGGCATTCGCGTGA	1820
Oy	1806	CTGGGGCAAGAC	CCCC	1821
Oy	1821	CGACGCAACGAC	CGCC	1836

LOCUS	DEFINITION	1781 bp	mRNA	linear	EST 03-JUN-2005
CX099623	RECM0260 A normalized whole-life-cycle cDNA library of rice <i>Oryza sativa</i> (indica-cultivar-group) cDNA clone E1057014, THREEB09, E1057004, E1062D05, E1072714, E1070G16, E108 5', mRNA sequence.				

ACCESSION	CX099623
VERSION	CX099623.1
KEYWORDS	GI:66912775
SOURCE	EST.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

REFERENCE
1 (bases 1 to 1781)

AUTHORS	TITLE
Zhang, J., Feng, Q., Jin, C., Qiu, D., Zhang, L., Xie, K., Yuan, D., Han, B., Zhang, Q. and Wang, S.	Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica

rice cultivar Minghui 63
Journal Plant J. 42 (5), 772-780 (2005)
PubMed 15918889
Comment Contact: Wang S

National Key Laboratory of Crop Genetic Improvement
Hauhung Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shiipingwang@hotmail.com
Seq primer: T7.

FEATURES	Location/Qualifiers
source	1..1781
	/organism="Oryza sativa (indica cultivar-group)"
	/mol_type="mRNA"
	/strain="indica"

```

/cultivar="Minghui 63"
/db xref="taxon:39946"
/clone="E1057014, THREEB09, E1057004, E1062D05, B1027114,
E1070G16, E108"
/tissue type="whole plant"
/dev stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_id="A normalized whole-life-cycle cDNA library of
rice"
/notes="Vector: pSPORI1; Site 1: SalI; Site 2: NotI; The
library is constructed based on the strategy of saturation
hybridization with genomic DNA using rice cultivar Minghui
63. This library consists of cDNA from 15 directionally
cloned cDNA libraries constructed with different tissues
from 9 developmental stages."

```

ORIGIN

Query Match 3.7%; Score 90.2; DB 8; Length 1761;
Best Local Similarity 43.8%; Pred. No. 6.5e-06;
Matches 447; Conservative 0; Mismatches 568; Indels 6; Gaps 1;

QY	231	GGA	CGTGGCCTT	CCCCAGGGCAAGGGCCCGCGAGTTCCCGACGAGCAAAACCGCGCAA	290	
Db	194	GGA	CGCGGTGTCCA	CGCGTGTCTCGCCCAAGAACCCCGACAGCAAGGTGGCTTGCAGAC	253	
QY	291	CAG	CCCCACCA	CGCGCGAGCTGCAGGTGCGCGGGAACAACCCCGACGAGGCGCGCGC	350	
Db	254	GTC	CAACCAAGACCA	AACTGGTGAATGGTGTTCGGGAGATCAACCAAGGCAACGTCGA	313	
QY	351	CGA	GCGCAGGGGCA	CCCTGAACTTCCCCAGATCAACCTGTGGGAGGAGCGCCCTCGTGGAG	410	
Db	314	CTA	GAGAAAGATCGT	CCGCGCAACTCTCCGCGGATCGGCTTCTGTCTCGACGACGTCGCG	373	
QY	411	CAT	CAGTGGGCGG	CCAGATCAAGAGAGGCCCTGTGGAACAACGCGGCGCGACGACCGT	470	
Db	374	CCT	GACGCCCA	CCGCTGCAAGGTGTCTGTAAATGACGACGATGCGCCGCAATTCGC	433	
QY	471	GCT	GAGAGAGATGA	AGCTGCCCCGCGCAAGTGAACCCAGATGATTCGCGCGCATTCGCGG	530	
Db	434	GCA	GGGGGTGCAC	GGGCTCACTTCAACAAAGCGCCCGAAGAGATCGGCGCGCGGACGAGG	493	
QY	531	CTT	CATCAAGGTG	CGCGCATGACACAGATCTTGATCGAGATCTGCGGCAAGAGCCAT	590	
Db	494	CCA	CACTGTTGGG	CTACCGCACCGACGAGACC	CGGAGTATGGCCCTCAGGCCACGTCCT	553
QY	591	CGG	CACCGTGTGAT	CGGCCCCACCCCGGTGAACATATCGGCGCGCAACATGCTGACCA	650	
Db	554	CGC	CAACCAAGGCT	CGGCGCGGCTCTCACAGAGTCCGCAAGAACGGGCACTGCGCGTGGCT	613	
QY	651	GCT	GGGCTGCAC	CCCTGAACTTCCCCATACGCCCATGAGACCGTGCCCGGTGAAGCTGAA	710	
Db	614	CAG	GCCCA	CGGCAAGACCCAGGTACCGGTGAGTACTTCAACACACGCGCGGCGCATGTGT	673	
QY	711	GCC	GGGCAATGA	CGGGCCCCAAGGTGAAGACAGTGGCCCCCTGAACGAGAGAAATCAAGGC	770	
Db	674	CCC	GTCGCGCTT	CAACCGTCTCATTTTCAACCAAGACGACGAGACCGTCAACAAAG	733	
QY	771	CCT	GACGCACTT	TCGAGAGATGAGAAAGAGGGCAAGATCAACAAATCGGCCCCGA	830	
Db	734	CGA	GATCGCCGCC	-----GACTCTAAGAGACAGTATCAAGCGGTCAATCCCGCAAA	787	
QY	831	GAA	CCCTTACA	CAACCCCGTGTTCGCCATCAAGAAAGAGACAGACCAAAAGTGGCGCA	890	
Db	788	GTA	CTCGACGA	GAAGAACCATCTTCTCAACTCAACCCCTCGGGCGCTTCGTCAATCGGCGG	847	
QY	891	GCT	GGTGAATTT	CGCGAGCTGAACAAAGCCGACCCAGGATTTCTGGGAGGTGCACTGGG	950	
Db	848	GCC	CAACGCA	CGCGCGCTCACCGGCGGCAAGATCATCTTCACACCTTACGCGCGAGTG	907	
QY	951	CAT	CCCCAC	CCCGCGGCTGAAGAAAGAAAGAGCGTGAACGTTGCTGAGACGTGGGCGA	1010	
Db	908	GGG	CGCGCA	CGCGCGCGCGCTTCTTCGGCAAGAACCCGACCAAGGTGCACGAGCGG	967	

Db	Accession	Version	KeyWords	Source	Organism
1011	CGCCTACTTCAACCGTGGCCCTTGAGACGAGACCTTCCGCAAGTACACACCGCTTACACATCCC				
1011	CGCCTACTTCAACCGTGGCCCTTGAGACGAGACCTTCCGCAAGTACACACCGCTTACACATCCC				
968	CGCCTACTTCAACCGTGGCCCTTGAGACGAGACCTTCCGCAAGTACACACCGCTTACACATCCC				
1071	CAGCATCAACACGAGACCGCCCGGATCCGCTACACGATCAACAGCTGCGCCCAAGGCTG				
1028	CATGTGTGAGAGTGTGTATCGCATCGCATTCGGGCTTCCCGAGCCGCTCTCCGTGTGTGTGACTC				
1131	GAAGGGCAGCCCGACGATCTTCCAGAGCAGATGACCAAGATCTCTGAGCCCTTCCGCGC				
1088	CTACCGGACCGGACGATCTCCCGACAGAGATCTCTCAAGATCTGTCAGAGAGAACTTTCGA				
1191	CCGACACCCCGAGATCGTATCTACAGAGCCCGCTGATCGTGGGACGAGCTGAGAT				
1148	TTTTCAGGCTCCGGAGATGATGATCATCACTGACTCAAGAGAGGCGGACACGGTTCAT				
1251	C 1251				
1208	C 1208				
RESULT 4	AY103647	2598 bp	mRNA	linear	HTC 18-FEB-2005
LOCUS	AY103647				
DEFINITION	Zea mays PC0142084 mRNA sequence.				
ACCESSION	AY103647				
VERSION	AY103647.1	GI:21206725			
KEYWORDS	HTC.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Bukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae, PACCAD clade, Panicoidae, Andropogoneae, Zea.				
AUTHORS	1 (bases 1 to 2598) Gandier, J., Schroeder, S., Polacco, M.L., Sanchez-Villada, H., Faring, Z., Morgante, M., Landwe, T., Fongler, K., Useche, F., Hanafey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C. and Coe, E.H. Jr.				
TITLE	Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization				
JOURNAL	Plant Physiol. 134 (4), 1317-1326 (2004)				
PUBMED	15020742				
REFERENCES	2 (bases 1 to 2598) Hainey, C.E., Dolan, M., Miao, G.H., Vogel, J.M., Whitelitt, M.S., Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.				
AUTHORS	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes				
TITLE	Unpublished (2002)				
JOURNAL	3 (bases 1 to 2598)				
REFERENCE	Coe, E.H.				
AUTHORS	Direct Submission				
TITLE	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA				
JOURNAL	Missouri, Columbia, MO 65211, USA				
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, MaizeMap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.				
FEATURES	Location/Qualifiers				
SOURCE	1..2598				
	/organism="Zea mays"				
	/mol_type="mRNA"				
	/db_xref="MaizeGDB:638378"				
	/db_xref="Taxon:4577"				
	/clone_lib="Maize Mapping Project/DuPont Cornsensus Library"				
	/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the				

ORIGIN

overgo addressing of BACs in conjunction with the Maize Mapping Project"

Query Match	Similarity	3.7#	Score 90.2	DB 4	Length 2599
Best Local	42.2#	Pred. No. 5.7e-06			
Matches 814	Conservative 0	Mismatches 1096	Indels 15	Gaps 5	
QY	296	CCACCAGCCGCGAGCTGCAAGTGGCCGCGCAACCCCGCAGCGAGGCCGAGCCGAGC	355		
Db	51	CCACCACCTCCCACTCTGCGCACCGCGCGGCCCAACCAACACACCGCAGCGCGACA	110		
QY	356	GCCAGGGCAACCTGAATTTCCCCAGATCACTCTGTGGCAACGCCCTCTGTGAGCATCA	415		
Db	111	TGGCGGGCAAGCGCGCCATCTGTGAGAGCGACCGCGTGAATCTGGGGCCCGCGCGCG	170		
QY	416	AGGTGGCGCGCAATCAAGAGGCCCTGTGTGACACCGCGCGCGACGACACCGTCTGG	475		
Db	171	AGCTGGCGGGAGCGCACTTGGACGAGGTGAAGCGCATGTGTGGCCGACAGGCCCGGACGCCG	230		
QY	476	AGGAGATGAGCTGGCCCGGCAATGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTCA	535		
Db	231	TGTCTAAGATGAGAGGTTCCACCTCTCGCGTGGCGCAGTGTGGCGCGCGCTCTCCCTCGCGCA	290		
QY	536	TCAAGGTGGCGCAATGACGACCAATCCGATCGAGATCTGTGGCGCAAGAAAGGCCATCGGCA	595		
Db	291	AGGACGGGTCGGCGGTCGCGGTGAGCTCGACGAGAGAGGCCCGCGCCCGGCTCAAGGCCA	350		
QY	596	CCGTGTGATCGGCCCCACCCCGGTGAACATCATCGCGCGGCAATGTCTGACCCAGCTGG	655		
Db	351	GCACGGAGTGAATCTTGACATGCGATCGCCACCGCGGGGCAATCTAACGGCGTCAACACG	410		
QY	656	GCTGACCCCTGAATCTTCCCCATCAGGCCCATTCGAGACCGTGCCTGTGAAGCTGAAGCCCG	715		
Db	411	GCTTCGGGGGCACTTCCACCGCGGCACCAAGGACGGGCGCGCTCCAGGATCGAGCTGC	470		
QY	716	GCATGGAACGGCCCCAAGGTGAAGAGTGGCCCTCTGACGAGAGAGAAATCAAGGCCCTGA	775		
Db	471	TCAGGCATCTCAAGCCCGGAATCTTGGCACCGGACGAGACGGGCAACGCTGCGTGG	530		
QY	776	CCGCGATCTCGAGAGATGGAAGAGGCAAGATCAACAAATCGGCCCCGAGAAAC	835		
Db	531	AGGTACACCGCGCGGCGATGCTGTGGCGCATCAACCTCTCCAGAGGCTTACTCCGACA	590		
QY	836	CCTAACAACCCCGGTGTTGGCCATCAAGAAAGAAAGACGACCAAGTGGGCGCAAGCTGG	895		
Db	591	TCCGTTTGAGATCTCGAGGCATTCAGAAAGCTGTCAACACCGGTGTCAAGCCCTGGCC	650		
QY	896	TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGAGTGCATCGGAGCATCC	955		
Db	651	TGCGGCTCCGGGGGACCATCAACCGGTGGGCGACGTGTCCGCTCTCTTACATCGCG	710		
QY	956	CCCAACCCCGCGGCTTGAAGAAAGAAAGACGTGACCTGTGCTGAGCTGTGGGCGACGCTT	1015		
Db	711	GCTTCATCACGGGCGGCCCAACGCGCAGGCGGTCAACG---TGACCGGAGGAAAGTGG	767		
QY	1016	ACTTCAGGTCGCCCTGAGCAGGACTTCCGCAATGACACCGGCTTCAACATCCACAGA	1075		
Db	768	ACGCGCGGAGGCGCTTCAAGATCGCGGCAATCGAGGGCGGCTTCTTCAAGCTCAACCCCA	827		
QY	1076	TCAACAACGAGACCCCGGCAATCGGCTAACAGTAAACAGTGTGCCCCAGGAGCTGAAGG	1135		
Db	828	AGGAGGGCGTGCATGTGTAACGGCACAGTCGATGGGCTCGCGCTCGCGCACCGTGA	887		
QY	1136	GCAAGCCCAAGATTTTCAGAGCAGCATGACCAAGATCTTGAAGCCTTCCGCGCGCGCA	1195		
Db	888	TGTACGACGCAAGTCTCTGGCGGTCTCTGTGGAGGTCT---GTCCGCGCTTCTTCTGGCG	944		
QY	1196	ACCCCGAGATGATCTAACAGAGGCCCGCTGTAGTGGGCGAGGCACTTGAAGATCGGCG	1255		
Db	945	AGGTCTATGAGCGGCAAGCCGAGTACAGGACACACTGACCCACAAGCTTGAAGCACACC	1000		
QY	1256	AGCACCGGCGCAAGATCGAGGAGCTGGCGACGACTGTG---CGTGGGGCTTCAACA	1312		

Db	1005	CGGGGTCCATCGAGGCCGCGGCCCATCATGAGGACATCTCTGGATAGGAGCTCTTCATGA	1064
Qy	1313	CCCCGACAGAGAGCAACAGAGAGGCCCTTCTCTGCCCATCGAGCTGACCCGACA	1372
Db	1065	AGCAGGCGCAAGAAATTGACAGAGCTGAGCCCGCTGCTGAGGCCCAACAGAGCAGGTACG	1124
Qy	1373	AGTGAACCGTGCAGCCCATGAGGTGCGCCGAGAGAGAGAGAGCTGAGACCGTGAACGACATCC	1432
Db	1125	CGCTTCGCACTGCGCCGACGAGGCTTGGGCCCCCAAGATCGAGTCAATCGCGCCGACCA	1184
Qy	1433	AGAACTGTGTGGGCAAGCTGAATCTGGGCAAGCCAGATCTACCCCGGACATCAAGGTGCGCC	1492
Db	1185	AGTCCATTCGAGCGGAGGGTCAACTCCGTAACGA---CAACCCGGTCAATCGACGTCCACC	1241
Qy	1493	AGCTGTGCAAGCTGTGCGCGCGCCCAAGGCCCTGACCGCAATGTGTCCCTTGACCCAGG	1552
Db	1242	GCGGCAAGGCGCTGACGCGCGGCACTTCAGGCGACCCCATGAGCGGTCTCATAGACA	1301
Qy	1553	AGGCGGAGCTGAGACTGTGGCCGAGAACCGCGAGATCTGTGCGGAGCCCGGTGCACGCGGTGT	1612
Db	1302	ACGCGCGGCTGCGCATGCGCCAAACATCGGCAAGCTCATATTTGGCGAATTCCTCGAGCTCG	1361
Qy	1613	ACTACGACCCCAAGAGAGCCTGTGTGCGGAGATTCAGAGACAGGCGCAAGCAATGTGA	1672
Db	1362	TCAACGAGTTCTTAACAACAACGCGGTCACTTCCAACCTGGCCGCGCAGCGCCCAACCCAGCC	1421
Qy	1673	CCTACCGAGATTACAGAGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCAAGATGC	1732
Db	1422	TGGACTACGCTTTCAGAGGSCACCGAGATGCGCATGTGCTCTTACTGTCTCGAGCTCCAGT	1481
Qy	1733	GCACCCGCCCAACCAACGACGTGAAGACGTGACCGAGGCGGTGCAGAGATGCGCATGG	1792
Db	1482	ACCTGGGCAACCCCATCAACAACGTCGAGACGCGCGAGACACACAACAGAGACGTGA	1541
Qy	1793	AGAGCAT---CGTATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATTCAGAAAGAGA	1849
Db	1542	ACTCCCTGTGGGCTGTGCTCGGSCAGAAAGACCGCGAGGCGATGACATCTCTGAAGCTCA	1601
Qy	1850	CCTGGAGAACCTGTGTGAACCGACTACTGTGCAGGCGCACTGTGATCCCGAGTGGGAGTTGG	1909
Db	1602	TGTGTGTCACCTAACATCGTGTGGCGCTGTGCCAGGCGGTGGAACCTGTGGCACTCGAGGAGA	1661
Qy	1910	TGAACACCCCTCCCTGTGTGAAGCTGTGTACAGCTGTGAGAGAGAGCCCATCATCGGCG	1969
Db	1662	ACATTCMAAGCGTGTGTGAAGAACACCGTGAACCGAGTGGCCAAAGAGGTGTGCACATGA	1721
Qy	1970	CCGAGACCTTTCACGTGAGCGGCGCGCCCAACCGGAGAACCAAGATCGGCAAGGCGGCT	2029
Db	1722	ACCCCTCGGGGAGAGTCTTCACAGCCCGCTTCAGGAGAGAGAGTGAATCAGCGGCATCG	1781
Qy	2030	ACGTGAACCGAACCGGGGCGGCAAGATCTGAAGCTTGAAGCCGAGACCAACAACAGAGA	2089
Db	1782	ACCGGAGGCGGTGTTCACGTACCGGGAGAGACGCGGCACAGGCGCAGCTGTGCGCTGAATGC	1841
Qy	2090	CCGAGCTGCAGGCCCATTCAGCTGTGGCTTCGACAGACAGGCGGACGAGAGGTGAACATCTGGA	2149
Db	1842	AGAACTCGGCGCGGTGTGTGTGAGACACGCGCTTACGACAGGCGCACCGGAGCGGGAGC	1901
Qy	2150	CCGACAGCCAGTACGCGCTCTGGGCATCATTCAGGCCCAAGCCGCAACAAGACGAGCGAGC	2209
Db	1902	CCTCGGTGTCTTCCAAGATCAACAGATTCAGAGAGAGACTCTCGGCGGTGTGCTGCCAGG	1961
Qy	2210	TGTGTGA 2216	
Db	1962	AGGTGA 1968	

RESULT 5	CL973991	LOCUS	DEFINITION
	CL973991	3069 bp	DNA
			linear
			GSS 21-SEP-2004
			OsIFCC025416 <i>Oryza sativa</i> Express Library <i>Oryza sativa</i> (indica cultivar-group) genomic, genomic survey sequence.

Accession	Version	Keywords	Source	Organism	Reference Authors	Title	Journal Comment	Features	Origin
CL973991	CL973991.1	GI:52402507	GSS.	Oryza sativa (indica cultivar-group)	Oryza sativa (indica cultivar-group)	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Wang, G. K. S., Deng, X. W. and Wang, J.	Department of Bioinformatics Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.	1.3069 Location/Qualifiers	
Query Match	3.6%	Score 88.8	DB 10	Length 3069					
Best Local Similarity	41.5%	Pred. No. 1.2e-05							
Matches	652	Conservative	0	Missmatches 917	Indels 3	Gaps 1			
37	ACGACGCGCAATCTCTGATCAGCGAGCACTTCAAGAGGCGCCAGAGCATCATCAAG	96							
625	AGCGGCATCGGCAGAGACGCGCTGGCGAGCTTACCGGAGGAAACAGATCCGAGAG	684							
97	TGCTTCAATCGCGGCAAGAGAGGCGCACATGCGCCGCAATGCGCGCGCCCGCGAAGAG	156							
685	CACCTTCGCTCAGCGGCGTGGTCCAGTCCCGCAAGTTCCGCTTCGCGACCTCTC	744							
157	GGCTGCTGAAGTGTGGGAGAGAGGCGCCACAGATGAAGAATGACACCGAGCGCAGGCG	216							
745	CACGACGTCAATCGCGCAGGCGCAAGCGCTCTCTCGGCGAGCAGCTGAGCGGAGGCG	803							
217	AACTTCTTTCGCGAGAGACTGGCGCTTCCCGGAGGAGAGCGCCGAGTTCCCGAGCAG	276							
804	--CGCCACAGACGACGTGCGCGCGCGGGGTGGATGAGAGCGGAGAGGCTGTCTAAGCG	861							
277	CAGAACCGCGCAACAGCCCGACAGCGCGAGCTGACAGTGTGCGCGGAGCAATCCCGCG	336							
862	CGCCTCGCGCGCGGCTCAGAGGAGAGGTAACTCTGAGTGTCTGACAGATCGGAGCG	921							
337	AGCGAGCGCGCGCGCGAGCGCGCAGGCGCCTTGAATTTCCCGCAGATCAACCTGTGAG	396							
922	AGGAGAGAGTGGGCTTCTTCTTCTGCGCGCTTCCCGCGGCGCTTGTGCGGAGCTGTG	981							
397	CGCCCGCTGTGAGCATCAAGTGTGGCGCGCAGATTCAGAGAGCGCTGTGAGACCGCGC	456							
982	CTCTGTACAGACGAGTCTAATATCAGCAGGAGTGTGAGCAACCTGTGCGCGCGCGCG	1041							
457	GCCGACGACACCGTGTGAGAGAGATGAGCTTGTCCCGGCAATGTGAAGCCCAAGTATC	516							
1042	GCGCTGTGCGCACACCGAGGAGCTGTGGAGAGCTGTGCGCGGAGAGCGCGCGCACTGTTC	1101							
517	GCGCGCATCGCGGCTTTCATCAAGTGTGCGAGTACGACGAGTCTGATCGAGTCTGCG	576							
1102	CGCGGAGAGGTGTATGCGTACCGGAGGCGGAGCGCGAGGAGGAGTGTGAGCAGCTCAGAGC	1161							

QY 577 GCGAAGAAAGCCATCGGCAACCGTCTGATTCGGCCCAACCCCGGTGAACATCATCGGCGCC 636
 Db 1162 CTCGGGAGCTCATGACGAAAGGGCTTCAACCTGCGCTCAACATGTCATGTCGCGGGA 1221
 QY 637 AACATGTCAGCCAGCTGGCTGACCTTGAACCTTCCCATTCAGCCCATGAGACCGTG 696
 Db 1222 CTCCTGAGGTCCAAAGAGAGATGATGAGGCGCTGATTCATAGCTTCATGATGATG 1281
 QY 697 CCCGTAAGCTGAAGCCCGGCAATGAGCGGCCCAAGTGAAGCACTGGCCCTGACGAG 756
 Db 1282 CCAACCCCGGCAATCGCCGGAAGAGACGCGCCGAGGCGAGCCGAAGAGACAGCTTC 1341
 QY 757 GAGAAATCAAGCCCTGATCCGCTCTGCGAGAGATGAGAGAGAGAGAGAGATCAAC 816
 Db 1342 GATGATACGAGACGAGCCCGGATCATGACGAGAGCAAGAGAGAGAGAGAGAGAG 1401
 QY 817 AAGATCGGCGCCGAGAGACCCCTTCAACACCCCGTGTTCGTCATCAAGAGAGAGAC 876
 Db 1402 CAGATGAGGACATTCATGAGCAAGATCTGACGAGTGTGATGAGACGACCTGCGCAG 1461
 QY 877 ACCAAGTGGCGCAAGCTGGTGAATTCGCGAGCTGAACAAAGGCAACCGAGACTTCTG 936
 Db 1462 CTGAAGCCGTGCTTCTTCTTCTTCTGCGCGGTTCAACCGGCAAGACCGCAATCTGCGCG 1521
 QY 937 GAGGTGACGCTGGGATCCCGCAACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAG 996
 Db 1522 AAGCTGTCCGGCTGTGGGTGCGCGAGGGTTTATGAGGCGGAGAGAGAGAGAGAG 1581
 QY 997 CTGAGACGTGGGAGAGGCTTCTTCAAGCTGCTTGAAGAGAGAGAGAGAGAGAGAG 1056
 Db 1582 GAGGAGCAACGCGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1641
 QY 1057 GCCTTCAACATCCCGAGATCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116
 Db 1642 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1701
 QY 1117 CTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1176
 Db 1702 TTCTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
 QY 1177 GAGGCTTTCGCGCGGCAACCCGAGATGATGATCAAGAGAGAGAGAGAGAGAGAG 1236
 Db 1762 CTCTCAACGCGCGCGCGCGCTGCGCTGCGCAACCTTCAAGAGAGAGAGAGAGAG 1821
 QY 1237 AGCGACCTGAGAGATGAGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
 Db 1822 GTCATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1881
 QY 1297 CGCTGGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
 Db 1882 GCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1941
 QY 1357 GAGCTGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1416
 Db 1942 CCCACCTTCTCGTCCAGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2001
 QY 1417 ACCGTAAGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
 Db 2002 CTGCGGCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2061
 QY 1477 GGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
 Db 2062 GGCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2121
 QY 1537 GTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596
 Db 2122 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2181
 QY 1597 CCGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
 Db 2182 CCTGCGCTCGC 2193

RESULT 6
 CL961989
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatoidae; Oryzaceae; Oryza.
 1 (bases 1 to 1398)
 REFERENCE
 AUTHORS
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Xiao, Y., Sun, N., Zhang, X., Bao, U., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W., and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 JOURNAL
 COMMENT
 Contact: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Location/Qualifiers
 1..1398
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"
 ORIGIN
 Query Match 3.5%; Score 86.6; DB 10; Length 1398;
 Best Local Similarity 45.9%; Pred. No. 2.8e-05;
 Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;
 QY 1 GTCCAGCCACATGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60
 Db 166 GTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 225
 QY 61 GCGAGCACTTCAAGAGGCGCCCAAGGAGATCAAGTGTCTTCACTGCGGCAAGAGAGG 120
 Db 226 GCGGCGGCTGCGGCGCGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 282
 QY 121 CACTGCGCGGCACTGCGCGGCGCGCGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAG 180
 Db 283 CGAGTATGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 342
 QY 181 GAGCAACGATGAAGAGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
 Db 343 GTCAAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
 QY 121 CACTGCGCGGCACTGCGCGGCGCGCGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAG 180
 Db 283 CGAGTATGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 342
 QY 181 GAGCAACGATGAAGAGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
 Db 343 GTCAAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 402
 QY 403 TTCCAGATCTCTCAACCCGAGAGTCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAG 462
 Db 463 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 513
 QY 463 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 513
 Db 513 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 513
 QY 361 GAGCACTGGAATTTCCCGAGATCACTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
 Db 421 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 480
 QY 514 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 570
 Db 571 ACACCAACCTGTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 630

QY 481 ATGAGCTGCTCCCGGCAAGTGAAGCCCAAGATGATCGGGCGATCGGCGCTTCATCAAG 540
 DB 631 AGGATCTGCTCGCTCAAGCCCAATGTCGGCTCTGGAGGCCACCGGCGCAAGCTCGCC 690
 QY 541 GTGCGCCAGTACGACCAAGATCTGATCGA---GATCTGGCGCAAGAGCCATCGGCAAC 597
 DB 691 GAGGCTGCTTACGCGCGTGGCAAGACTTCAGAGATCGGCGCAAGTTTCTCAACCGCAGC 750
 QY 598 GTGCTGATCGGCGCCACCGCGTGAACATCATCGGCGCAAGATCTGACCGAGCTGGCC 657
 DB 751 GTGGGATTCGGTGGCTCTGCTTCCAGAAAGCATCTCACTGCTG----- 798
 QY 658 TGCACCTTGAATCTTCCCATCAAGCCCATCGAGACCTGTCCTGTAAGCTGAAGCCGAGC 717
 DB 799 TACATCTGCGAGTGCATAGGCTCTCCCGAGGCGCAACTACTGGAAGCAAGTTATCAAG 858
 QY 718 ATGGAAGGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGAGAAATCAAGGCGCTGACC 777
 DB 859 ATCAACGACTTACCAAGAGCGCGTTCTGTAACCGCGTGTGTCTCATGTTCAACAGCC 918
 QY 778 GCCATCTGCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATCGGCGCCGAGAACCC 837
 DB 919 GTGGCGCGCAAGAGATGCGCGTGGCTTCCGCTTCAAGAGAGACCGGCGACACC 978
 QY 838 TACACAACCCCGCTGTGTTCCCATCAAGAGAGAGAGACCAAGTGGCGCAAGCTGTG 897
 DB 979 AAGGAGACCGCGCCCATCAAGCTGTCGAAGGCGCTGATCGCGCAAGGCGCAAGGTGAC 1038
 QY 898 GACTTCGCGAGGTGAAGAGGCAACCGCAAGCTTCTGGAGGTGAGCTGGGCAATCCC 957
 DB 1039 ATCTACGATCCCCAGGTGACGAGAGACAGGTCTCAAGCGGACCTCGCCATGAGCAAGTTC 1098
 QY 958 CACCCCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1017
 DB 1099 GATTTGGATCAACCGGCTCATCTCCAGCGAGTGAAGCCCAACGCGCATCAAGAGGTGTC 1158
 QY 1018 TTCAAGGTGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1077
 DB 1159 GTGCGCTTGGAGAGCGCTTCAAGAGGCGCGCAAGGAGCGCCACGCGTCTGCACTCCAG 1218
 QY 1078 AACACGAGAGACCGCGCGCATCGCTAC 1104
 DB 1219 TGGAGAGATTCAAGAGCTTCACTAC 1245

RESULT 7
 C2247380 743 bp DNA linear GSS 11-FEB-2005
 LOCUS A1AA-aat35a12.b1 Ancylostoma caninum whole genome shotgun library
 DEFINITION (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
 ACCESSION C2247380
 VERSION C2247380.1 GI:59632821
 KEYWORDS GSS.
 SOURCE Ancylostoma caninum (dog hookworm)
 ORGANISM Ancylostoma caninum
 Buxley, J.; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 1 (bases 1 to 743)
 Mitreva, M., McCarter, J.P., Pape, D., Rittler, R., Tsagaris, I., R.,
 Ronko, I., Martin, J., Wylie, T., Danie, M., Meyer, R., Messina, D.,
 Waterston, R.H., Clifton, S.W., and Wilson, R.
 Genome Survey sequences from the parasitic nematode Ancylostoma
 caninum
 Unpublished (2004)
 Contact: Mitreva, M
 Washington University in St. Louis
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: nematode@watson.wustl.edu
 Genomic DNA provided by John Hawdon (mtmjnh@wumc.edu) DNA

Sequenced by Washington University Genome Sequencing Center
 Class: shotgun.
 Location/Qualifiers
 1..743
 /organism="Ancylostoma caninum"
 /mol_type="genomic DNA"
 /strain="Baltimore"
 /db_xref="taxon:29170"
 /dev_stage="Adult"
 /lab_host="GS10"
 /clone_lib="Ancylostoma caninum whole genome shotgun
 library (A1AAGSS 001)"
 /note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
 Ancylostoma caninum genomic DNA was randomly sheared,
 end-repaired and size fractionated to enrich for 2-4 kb
 fragments. Genomic DNA was provided by John Hawdon
 (mtmjnh@wumc.edu) at George Washington University.
 Sequencing by Washington University Genome Sequencing
 Center, St. Louis, MO."

ORIGIN
 Query Match 3.5%; Score 86; DB 10; Length 743;
 Best Local Similarity 45.1%; Pred. No. 3.5e-05;
 Matches 320; Conservative 0; Mismatches 390; Indels 0; Gaps 0;
 777 CCGCATCTGCGAGAGATGAGAGAGGCGCAAGTCAACCAAGATCGGCGCCGAGAACCC 836
 DB 5 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 64
 QY 837 CTACAAACACCCCGTGTGCTCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
 DB 65 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 124
 QY 897 GGAATTCGCGAGCTGAACAGCGGACCCAGAGCTTTGGAGGTGAGCTGGGCAATCCC 956
 DB 125 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 184
 QY 957 CACCCCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1016
 DB 185 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 244
 QY 1017 CTTGAGCGGCGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1076
 DB 245 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 304
 QY 1077 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1136
 DB 305 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 364
 QY 1137 CAGCCCGAGATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1196
 DB 365 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 424
 QY 1197 CCGCGAGATCGATCTACAGGCGCCCTGTACGTGGGAGAGAGAGAGAGAGAGAGAGAG 1256
 DB 425 CAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 484
 QY 1257 GCACCGCGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1316
 DB 485 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 544
 QY 1317 CGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1376
 DB 545 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 604
 QY 1377 GACCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1436
 DB 605 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 664
 QY 1437 GCTGTGGGCAAGCTGAAGCTGGGCGCAAGATCTTACCCGCGATCAAG 1486
 DB 665 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 714

RESULT 8
BM321451 951 bp mRNA linear EST 03-JAN-2002
LOCUS BM321451
DEFINITION Mestigamoeba balamuthi lambda ZAP II library
Mestigamoeba balamuthi cDNA similar to adenomythomocysteinease (EC 3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857
KEYWORDS EST
SOURCE Mestigamoeba balamuthi
ORGANISM Mestigamoeba balamuthi
Eukaryota; Pelobiontida; Mestigamoebidae; Mestigamoeba.
REFERENCE 1 (bases 1 to 951)
Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Senses, C. W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., and Philippe, H.
The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mestigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
JOURNAL 11830664
PUBMED
COMMENT Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 951 Std Error: 0.00
POLYA-No.
FEATURES
source Location/Qualifiers
1..951
/organism="Mestigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_id="Mestigamoeba balamuthi lambda ZAP II library"
/note="syn: Phreatamoeba balamuthi"
ORIGIN
Query Match 3.5%; Score 85.6; DB 3; Length 951;
Best Local Similarity 46.0%; Pred. No. 4,2e-05;
Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;
QY 1722 CGCCAAAGTGCACCGCCGACCAACCAAGCTGAAGCAGTGAAGCCGCTGACAGAA 1781
DB 219 CGCCAAAGTGCACCGCTGTGTCTGTGCAACATCTTCGACGAGACCAACGCGCCGCGC 278
QY 1782 GATCGCCATGAGAGAGATGATGATGCGGACAGACCCCAAGTTCCGCTGCCATCCA 1841
DB 279 CATCCGCGAGCGCGCGCTCGCTCGCTGCGAAGGCGAGAACTTCAGAGATCATG 338
QY 1842 GAAAGAGACTTGGGAGACTGTGTGACCGAATGAGAGGCGCACTGTGATCCCGAGTG 1901
DB 339 GAGGTGACCTGAAAGGCTGTGTGCTTCGCGCCCTTACCAAGGCGCTTCAGATCATGTGCA 398
QY 1902 GAGGTTCGTGAACACCCCGCTGTGTAAGCTGTGATCACTGAGTGAAGAAAGAGCCCT 1961
DB 399 CGACGCGGTGACGAGACTGTGATGATCAAGAGGTTCCGCGCGAGAGCAACCCCA 458
QY 1962 CATCGCGCGGAGACTTCTTACGTGACGCGCGCCCAACCGCGAGACCAAGATCGGCA 2021
DB 459 GCTGTGAGAGACGACGAGAGGCTCGAGAGAGTCCGCTGCTCAACAGTCTCAAGCA 518
QY 2022 GCGCGCGCTACGTGACCGACCGGGGCGGAGAAATGCTGAGCTGACCGAGACCAACCA 2081
DB 519 GGTTCAGAAAGAGAGCGCGCGCTTCTGGACAAAGATCTCCCGGAGATCCGCGGTGCA 578
QY 2082 CCAGAGACCGAGCTGACAGGCGCATCCAGCTGCGCTGACAGACGAGCGAGAGGTGA 2141
DB 579 CGAGAGAGAGAGAGAGCTGCTTAAGAGCTGTACACAGCTGACCGAGAGAGAGCTGCT 638
QY 2142 CATGTGACCGACAGCGAGTACCGCTTGGGATCATCCAGCCGAGCCGAGCAAGAGCA 2201
DB 639 GTTCCCGGCGCTCAAGTCAAGAC--TCTTTCACCAAGAGCAAGTTGACAAATCTTA 695

QY 2202 GAGCGAGCTGTGAACCAATCATCGAGCAGCTGATCAAGAAGAGATGTAAGTCTGAG 2261
DB 696 CGGCTGCGCCCACTCGCTCATCGACGATCAAGGCGGACCGACCGAGTGAATGCTGGCGG 755
QY 2262 CTGGGTGCGCCCGCCCAAGAGGCGATCGCGCGGACAGAGAGATCGACAAAGCTGTGAGCA 2321
DB 756 CAAGGTGCGCGCTCGTCCGCGGCTACGCGAGATGTGGGCAAGAGGCTGCGCGAGTGTGCG 815
QY 2322 GCGCATCCGCAAGGTGCTGTTCTGAGACGATGATGCGGCGATGTGATCAAGTA 2381
DB 816 CGGCGAGGCTGCGCGCGCTGATCGACGAGATCGACCCCATCTGCGCGCTGACAGCGTC 875
QY 2382 CATGAGACGCTGTGATCGTGGGACGCGCGCGCTTACAGATCGA 2423
DB 876 GATGCGCGCTTCAGAGTCAACAGCTGAGGCGGCGCTCGA 917
RESULT 9
CL971508 1941 bp DNA linear GSS 21-SEP-2004
LOCUS CL971508
DEFINITION OsIFCC021485 Oryza sativa Expressed Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION CL971508
VERSION CL971508.1 GI:52397596
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1941)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W., and Wang, J.
An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
Unpublished (2004)
JOURNAL Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-8048676
Email: chenchen@genomics.org.cn
COMMENT Rice genomic sequence.
Class: exon-trapped.
FEATURES
source Location/Qualifiers
1..1941
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_id="Oryza sativa Expressed Library"
/note="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match 3.4%; Score 84.2; DB 10; Length 1941;
Best Local Similarity 42.5%; Pred. No. 7.8e-05;
Matches 705; Conservative 0; Mismatches 933; Indels 21; Gaps 4;
QY 1 GTGACGCGCACCATGAGCGGCGCATGAGCGGCGACCAAGCGCGCAATCTGATGAG 60
DB 163 GAGCGCGCCAGAGACGAGTGGCATGACCAACCAACCGTCTTGTGATGCGAAGCGG 222
QY 61 CGACGCACTTCAAGGCGCCCAAGGCGATCATCAAGTCTTCAACTGCGGCAAGAGGCG 120
DB 223 TTGATGTGTAAGAGTTCTCCGACCGCTCGGAGAGATGACATGAAGCTAATGCGCTTC 282
QY 121 CACATCGCGCGCACTGCGCGCGCCCGCGCAAGAGGCTGTGAAATGCGGCAAGAG 180
DB 283 AAGGTGTGCTGCGCGCGCGGAGCAAGCGCGATGATGTGTCAGTACAAAGGCGAAGAG 342
QY 181 GCGCACGAGATGAAGACTGACCGAGGCGCAAGCGCAACTCTTCGCGAGAGACTGCGC 240

Db 287 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 228

QY 1180 CCCTTCGCGCGCCGCAACCCCGAGATCGATCTACAGGCCCCCTGTACTGGGCAAC 1239

Db 227 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 168

QY 1240 GACCTGGAGATGGGCAAC 1258

Db 167 AACGGCAACAACAACAACA 149

RESULT 12
CZ216254

LOCUS CZ216254 892 bp DNA linear GSS 10-FEB-2005

DEFINITION A1AA-aat23g02.b1 Ancylostoma caninum whole genome shotgun library (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.

ACCESSION CZ216254

VERSION CZ216254.1 GI:59229909

KEYWORDS GSS.

SOURCE Ancylostoma caninum (dog hookworm)

ORGANISM Ancylostoma caninum
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida; Ancylostomatoidea; Ancylostomatidae; Ancylostomatinae; Ancylostoma.

REFERENCE 1 (bases 1 to 892)
Mitrev, M., McCarter, J.P., Pape, D., Ritter, R., Tsagaris, R., Ronko, L., Martin, J., Wylie, T., Dante, M., Meyer, R., Messina, D., Waterston, R.H., Clifton, S.W. and Wilson, R.
Genome Survey sequences from the parasitic nematode Ancylostoma caninum
Unpublished (2004)

TITLE Contact: Mitrev, M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@watson.wustl.edu
Genomic DNA provided by John Hawdon (mcmjmgw@umc.edu) DNA sequenced by Washington University Genome Sequencing Center
Class: shotgun.

FEATURES
source 1. 892
Location/Qualifiers
/organism="Ancylostoma caninum"
/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GSI0"
/clone_lib="Ancylostoma caninum whole genome shotgun library (A1AAGSS 001)"
/note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI; Ancylostoma caninum genomic DNA was randomly sheared, end-repaired and size fractionated to enrich for 2-4 kb fragments. Genomic DNA was provided by John Hawdon (mcmjmgw@umc.edu) at George Washington University. Sequencing by Washington University Genome Sequencing Center, St. Louis, MO."

ORIGIN

Query Match 3.3%; Score 80.6; DB 10; Length 892;
Best Local Similarity 46.5%; Pred. No. 0.00033;
Matches 334; Conservative 0; Mismatches 379; Indels 6; Gaps 2;

QY 1644 GATTCAGAGAGAGGCGGACGAGCTGACCTACGAGTCTACAGGAGCCCTTCAAGA 1703

Db 5 GATTCAGCTCGAGCACTTCAAGAGAGCTCAACGACGCTACCACTCTGACGT 64

QY 1704 CCTGAAGACCGCAAGTACGCCAAGATGCGACCCGCCACACCAAGAGCTGAAGAGCT 1763

Db 65 CCAAGCACAACAAGCAACAACATCCACAACAACCTTCAAGACACGACGACTAATGACT 124

QY 1764 GACCGAGGCGGTGAGAAATGCGCATGAGAGCACTCGTGTGGGGCAAGCCCCCA 1823

Db 125 CTAGACGCTCAAGCAGACGCTCAACGACAGCTCAACGACATCGAGCTGAGACAC 184

QY 1824 GTTCCGCTGCCCATCCAGAAAGAGACTGGAGACTGGTGAACCGACTACTGCGAGCG 1883

Db 185 CAGACCAATTAAGCTCCACACACCTTCAAGCAGACCGACGACTACTGACTTACACGT 244

QY 1884 CACCTGATCCCGAGTGGAGTTCGTGAACACCCCCCTGGTGAAGCTGTGTACCA 1943

Db 245 CAAAGCAGACGTCACCAAGCAGACGTCACAGACGTCACGACCAACCAACCCCA 304

QY 1944 GCTGAGAGAGAGCCCATCATCGGCGCGAGACTTCTACGTGAGAGCGGCCCAACCG 2003

Db 305 CCAAGCTCCACAACCACTTCAAGCAGACCGACGACTACTGAGTACCTGACACAGA 364

QY 2004 CGAGACCAAGATCGCAAGCGCGGCTACGTGACCGAGCGGGCGGCAAGATGTGAG 2063

Db 365 CGTCAACGACGACCTCGAGCTGACGACCAACCAACCAACGTCACCAACCTTCA 424

QY 2064 CTTGACCGAGACACCA---CGAGAACCGAGCTCGAGGCTATCGAGCTGGCTGCA 2120

Db 425 CGACACCGACGACTACTGCTCGACGACGTCACAGCAGCAGCTCAACGCTCAACCA 484

QY 2121 GAGACGCGGCGAGCGAGTGAAATGTCAGACGACGACGACGCTGAGGCTATGCA 2180

Db 485 GCACTCTCAAGCTCGACGACACCAAGACCAACGCTCAACCACTTCAAGACACCA 544

QY 2181 GGCACGCGCGGCAAGACGAGAGGAGCTGTGTAACCAAGATCATGAGCAGCTGATCA 2240

Db 545 CGACTACTAGCTCGACGACGTCACGACGTCACGACGTCACGACGACATGCA 604

QY 2241 GAAAGAGAGGTGTACTGAGCTGGTGGCCGCCCAAGAGGATCGCGCAAGACA 2300

Db 605 CGTCAACGACCAACAAGCAACAACGATCAACCAACGACGACGACGACGACTACTA 664

QY 2301 GATCGAAGAGCTGTGAGCAAGGAGGATCGGACGAGGTGTCTGTCGAGCGGATGAG 2359

Db 665 GCTCG---AGCAGTCAACGACGACGTCACCAACGACCTCGAGCTGACGACGAGG 720

RESULT 13
BM320864 1132 bp mRNA linear EST 03-JAN-2002
BM320864
LOCUS BM320864
DEFINITION rocketfeller.0.46 Mastigamoeba balamuthi lambda ZAP II Library
Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA sequence.

ACCESSION BM320864

VERSION BM320864.1 GI:18055270

KEYWORDS EST.

SOURCE Mastigamoeba balamuthi

ORGANISM Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE 1 (bases 1 to 1132)
Bapteste, R., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and Philippe, H.
The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

TITLE Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 1132 Std Error: 0.00
POLYA-No.

FEATURES
source 1. 1132
Location/Qualifiers
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Mastigamoeba balamuthi lambda ZAP II Library"

/note="syn: Phreatamoeba balamuthi"

Query Match	3.3%	Score 80.2;	DB 3;	Length 1132;
Best Local Similarity	45.3%	Pred. No. 0.00039;		
Matches 375; Conservative	0;	Mismatches 444;	Indels 9;	Gaps 2;

OY	139	CGCGCCCCCGCAAGAGGGCTGCTGAAAGTGGCGCAAGAGGGCGACCAATGAAAGAC	198
Db	79	CGCGCGAGGGCGAAGCGAATTACCGCGCGCGCCACTGGTGTATCGAAGACAGAACAG	138
OY	199	TGCACGAGCGCGCAGGCCACTTCTTCGCGAGAGACTGGCTTCTCCCAAGGCGAAGCC	258
Db	139	TACCAAGAGCCCAAGTACCGCTTGGTC-----GTCCGGTTCAACAAAGGAGCATCGTC	192
OY	259	CGCGAGTTCCCGAGGAGACGAACCGCGCCAAAGCCCCACAGCGCGAGCTTGAGGTG	318
Db	193	TGCCAGATCGCTTACGCCAAGATCGACGCGACCAATCTTCGCGCGCGCTTACTCGCAC	252
OY	319	CGCGCGCAACACCCCGCAGCGAGGCGCGGCGCGAGCGCGAAGGGACCTGTGAATTTCCC	378
Db	253	GAGCTCACCCGTTTGGGGGTCAAGCTTGGCTCTGACCACTACGCGCGGCTTACCGCACT	312
OY	379	CAGATCAACCTGTGACAGCGCCCCCTGTGTAGACATCAAGTGTGGCGGCGACATCAAGAG	438
Db	313	GCGCTGCTGTGGCGCCGCGCGTGTGTGAAGAAAGCTCAACTGTGACTCCAAAGTAAAGGGT	372
OY	439	GCCCTGTGACACCGCGCGCGACACACCTGTCTGTGAGAGATGAGCTTGC CGGCAAG	498
Db	373	GTCAAGAAAGGTCAAGGCGGAGGACTCAACGTGAGAGCTTCAGAGCGGCGCCCGCGT	432
OY	499	TGAAAGCCCAAGATATGATCGGCGGCGCATCGCGGCTTCATCAAGAGTGCACAGTAAAGAC	558
Db	433	TTCAAGGCGCTGTGAGCTGCGGCTGTGTGCGCACTGAGCTGGCGGCGCGGTTCGCG	492
OY	559	ATCTGATCGAGATCTGCGGCGAAGAGGCCATCGGACCTGTGTATCGGCGCCCAACCCC	618
Db	493	GCCCTCAAGGGGATGTGCGACGCGCGGCGTCAACGTCTCC---CAGAGCGAAGACCGGCTTC	549
OY	619	GTGAACATCATCGCGCGCAACAATGTGAACCCAGCTGGGCTGACCTTAACTTCCCCATC	678
Db	550	GTGCGCTTCAACGCGGACAGAAAGAGCTCAACGCGCGGCTCTTCGCAAGTCAATCTTC	609
OY	679	AGCCCATCGAGACCGTGCCCGCTGTGAAGCTGAAGCCCGGCAATGAAAGCGCGCCCAAGGTGAAG	738
Db	610	GCGCGCCACTGTGCGCGCGCTACATGAAGCTCTCAAGAGAGAGAGACGCGCGCGCTTCGAC	669
OY	739	CAGTGGCCCTTGACCGAGAGAAATCAAGGCCCTGAACCGGCATCTGTGCGAGAGATGAG	798
Db	670	CGCCAGTTCTCGCGCTACGCGCAAGAGAGGGTGTCAACGCGCAATGCTCGAAGAAATCTAC	729
OY	799	AAAGGCGCAAGATCAACCAAGTTCGGCCCCCGAAGAACCTTACAAACACCCCGTGTTCGC	858
Db	730	ACCGAGGCGCCACAAGCAAGATTCGCGGCCAACCGGACCTTGTCTCCCAAGGCGGCTCTGAG	789
OY	859	ATCAAGAAAGAGACGACCAAGTGTGCGCAAGCTGTGTGAATTTCCGCGAGCTGAAAG	918
Db	790	CCCGAGGGGGCCAAACCGCAAGCACTGGGCGCAAGCGCAGCTGACGTACCAACAGAGCGAAG	849
OY	919	CGCACTCCAGAACTTGTGGAGGTGCACTGGGCAATCCCCACCCCGCC	966
Db	850	AACCGCTGTGCGCAGAAAGGTTCGCGCTGTGCTACCCCGAAGCGCCCC	897

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PIBMED
COMMENT
FEATURES

Mastigamoeba balamuthi
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
1 (bases 1 to 1165)
Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,
Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.
The analysis of 100 genes supports the grouping of three highly
divergent amoebae: Dictyostellium, Entamoeba, and Mastigamoeba
Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
11830684
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockvax.rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA=No.

location/Qualifiers
1. 1165
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_11b="Mastigamoeba balamuthi lambda ZAP II Library"
/note="syn: Phreatamoeba balamuthi"

Query Match	3.2%;	Score 79.6;	DB 3;	Length 1165;
Best Local Similarity	45.1%;	Pred. No. 0.0005;		
Matches 428; Conservative	0;	Mismatches 506;	Indels 16;	Gaps 3

QY 90 CATCAAGTGTCTTCAACTGCGGCAAGAGAGGGCCACATGCGCCGCACTGTCTGCGCC 14.9

Db 46 CGTCAAGAAACAAGGCGTACTTTCAACGCGCTTCCAGACCAGTTCCGTGCGCGCGCAAGG 10.5

QY 150 CAAGAAGGAGTCTCTGAAAGTGCAGCAAGAGAGGGCCACAGATGAAAGACTGCACCAAGCG 20.9

Db 106 CAAGACGACCTACCGCGCGCGCCACNTGGTGTATCCAGACCAAGAACAGATGACACAGCCC 15.5

QY 210 CCAGGCCAACTTCTTCCGCAAGGACTGTGCGCTTCCCGCAGGCAAGGCCCGCAAGTTCC 26.5

Db 166 CAAGTACCGCTCTGTC-----GTCCGCTTTCACCAACAGGACATGTCGTCGCAGATCGC 21.9

QY 270 CAGCGACGAGAACCCCGCCCAACAGCCCAACAGCGCGAGCTGCAGTGTGCGCGCCACAA 32.9

Db 220 CTACCCCAAGATCGACCGGACCAATCTCTGCGCGCGCTTACTGCACAGACTCACCG 27.9

QY 330 CCCCACAGCGAGCGCGCGCGCGAGCGCCAGGGCACCTTGAACTTCCCCAGATCACCT 38.9

Db 280 CTTCCGCGCTCAAGCTCTGCGCTCAACCACTACCGCGCGCTACGCGACTGCGCTGCT 33.9

QY 390 GTGGCAGCGCCCCCTTGCTGAGCATCAAGGTGGCGGCGCAGATCAAGAGGCGCTGCTGA 44.9

Db 340 GCGCGCGCGTGTGCTGAAGAACTCAACTCTCACTCAAGTACGAGGGTGTCAAGAAAGT 39.9

QY 450 CACCGGCGCGCAACGACACCGTCTGGAGAGATGAGCGTGC CGCGGCAAGTGAAGGCCAA 50.9

Db 400 CAACGCGCAGAGCTACACATGTGAAGAGCTCGACGACGGGCGCCCGCGTTCAAGGCC-- 45.7

QY 510 GATGATCGGCGGACATCGGCGGCTTATCAAGGTGCGCCAGTACGACCAAGATCTGTGA 56.9

Db 458 --TGCTCGACGTGTGGCTGTGTTCGACCTCGACTGGCGCGCGCGGTGTTCGCGCGCCCAA 51.5

QY 570 GATTCGCGCAAGAAAGCCATGCGCACCGTGTGATCGGCGCCACCCCGTGAACTATCAT 62.9

Db 516 GGGCATGTGCGACGCGCGGCGTCAACGTCCGCCCAACGCGAGACCCGCTTGTGTGCGCTTCAA 57.5

QY 630 CGGCGCGCAACATGCTGACCCAGCTGGGCTGCAACCTGAACCTTCCCATCAGGCCCATCGA 68.9

Db 576 CGGCGACAGAAAGAGCTTAACGCGCGGATCTCTCGCAATGATCATTTTGCGCGCACGT 63.5

QY 690 GACCGTCCCGTGAAGCTGAGACCCGCGCATGACCGGCCCAAGTGAAGCATGTGCGCCT 74.9

Db 636 CGCCGCTGATGATGAGCTCTCTCAAGAGAGAGACCCCGCGCTTTCGACCCGCAATTCTC 695
Qy 750 GACCGAGAGAGAGATCAAGGCGCTTGAACCGCATCTGCGAGAGATGAGAGAGAGGCGCA 809
Db 696 GCGCTAACGCAAGAGAGGAGTGTACCGCGGACATGTCTGAGAGATCTACACCGAGGCCCA 755
Qy 810 GATCAACCAAGATCGGCGCCCGAGAACCCCTACCAACACCCCGTGTTCGCGCATCAAGAGAA 869
Db 756 -----CAAGCAGATCCGCGCGGACCGGACCTTTCGTCCTCCCAAGCGCGCTCGAAGCCCGA 809
Qy 870 GGACAGCAACCAAGTGGCGGACGCTGATGACTTTCGCGAGCTGAACAAGGACCCAGCA 929
Db 810 GGGCGCCAGGCGCCCAAGCACTGAGGCGAGCGGCTGACGTACACAGAGCGCAAGAACCG 869
Qy 930 CTTCTGAGAGTGAAGCTGAGCGATCCCGACCCCGCGCGCTGAAGAGAGAGAGCGCT 989
Db 870 CGTCCGCGCAAGAGAGAGTCCGCGTGGGCTAACCCCGACCGCCCAAGAGCGCAATAATTCC 929
Qy 990 GACCGTCTGAGACGTGGGCGAGCGCTTACTTTCAGCGTGGCGCTGAGAGAGG 1039
Db 930 CGGTCCCTGTGACACCGCGCATCTCCGCTGCGCGCTGCGCTGCTGCGG 979

RESULT 15
CL982770 1680 bp DNA linear GSS 21-SEP-2004
LOCUS Oa1PSC049024 Oryza sativa Expressed library Oryza sativa (indica
DEFINITION CL982770 genomic, genomic survey sequence.
ACCESSION CL982770
VERSION CL982770.1 GI:52420015
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group) Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrharioideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1680)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
Unpublished (2004)
JOURNAL Contact: Chen Chen
COMMENT Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

FEATURES
source Location/Qualifiers
1..1680
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 3.2%; Score 79.6; DB 10; Length 1680;
Best Local Similarity 41.9%; Pred. No. 0.00051;
Matches 612; Conservative 0; Mismatches 844; Indels 6; Gaps 2;

Qy 270 CAGCGAGCAGAACCGCGCCCAACAGCGCCGAGCTGCAAGTGGCGGCGCA 329
Db 129 CATCAACGTGAGACCGCGGACCACTGTCCCGCTTCAGCAAGGAGTGTTCGTACCGA 188
Qy 330 CCCCCGAGAGAGCGGCGCGCGAGCGCGCAAGGCGACCTTGAACTTCCCAAGATCAACCT 389
Db 189 CAGCGCGCGCGAGACCGAAGCTTCACTTCCGCGCACTACGAGCGTTTATGAAACGCGCAT 248

Qy 390 GTGGAGCGCCCGCTGTGAGCATCAAGTGGGCGCCAGATCAAGAGAGCGCTTCTGGA 449
Db 249 GAAGAGACCAACAACTTACACAGCGCGCGCATTAACAAGTCCGTGTGAGAGAGAGCG 308
Qy 450 CACCGCGCGCGAGCAACCGTGTGAGAGAGATGAGCTTCCCGCGCAAGTGAAGCCCA 509
Db 309 CCGCGCGCACTAACCTTCGCGCAAGAGCGGTGCAAGTATCCCGGACGTCAACAGAGATCA 368
Qy 510 GATGATCGGC---GGCATCGCGGCTTCATCAAGTGGCGCAATACAGATCTGAT 566
Db 369 GGAATACGTCAAGCGCGCGCGGCGATCGGACCGAGCGAGCGGTGAGAGTGGCATGCT 428
Qy 567 CGAATCTGCGCGCAAGAGCGCATATGCGACCGTGTATCGGCGCCACCCCGTGAACAT 626
Db 429 CGAGATCGCGCGCAACCGGTGCGAGATGGAATCGCTGCGCTTCGAGAGCGGTGCGCA 488
Qy 627 CATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTGAACCTTCCCATAGCGCCAT 686
Db 489 GATGAGCTTGGCGCATGGGCGCGAACATTCGCGCTTCGTGACCTCACTACCTGCGCTTA 548
Qy 687 CGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGAGCGAGCGAGCGCCAGAGTGAAGTGGCC 746
Db 549 CATCGCGCGCGCGGTGAATCAAGACCAAGCCACCGAGCAACAGGTGAGAGAGCTGCG 608
Qy 747 CCGTACCGAGAGAGAGATCAAGGCGCTTGAACCGGCATCTGCGAGAGATGAGAGAGAGG 806
Db 609 CGAGATTCGCGCATTCAGCGCGCGCGCATCTGCTGCGCGCGCGAGCGCGCGATCCGAGCA 668
Qy 807 CAAGATCAACAGATCGGCGCGCGCGAGAACCCCTTACCAACACCCCGGTTCGCGCATCAAGA 866
Db 669 GAGCGCGCGAGATCTCGCTGTTCACCAAGTGGCGCAATGAGGCGGTGATGAGATGAG 728
Qy 867 GAAGGACAGCAACCAAGTGGCGCGAGCTGTGAGCTTTCGCGAGCTGAACAGAGCAACCA 926
Db 729 GAGCGTGAACACCAATCTCAAGAGTGCCTCCGCAATGTGACGAGCGAGGCGTGAAGCGCT 788
Qy 927 GAGCTTCTGAGAGTGTGACGTGAGTGGGATATCCCAACCCCGCGCGCTGAAGAGAGAGAG 986
Db 789 GATCTGCGCAACAGCTGCGCGCTGAAGACCGCGCGCGCGCTCAAGCGCTGAGACCAAGT 848
Qy 987 CGTGAACGCTGAGCGTGGGCGAGCGCTTACTTGAAGCGCGCGCTGAGACGAGATCTTCCG 1046
Db 849 CGTGAACAGACCGGAGCATTCGCGAGGCGAGGCTCACTCGCATGATGAGGCAAGTACCT 908
Qy 1047 CAAGTACACCGGCTTACCATTCGCGAGATCAACAGAGAACCCCGCGCATTCGTAACA 1106
Db 909 AGACCTGTCCAGACACTACAGTCCGTGAGAGAGGCGGTGCGCGCGCGCATGAGAGA 968
Qy 1107 GTACAACTGTGCTGCGCGAGGCTGAGAGGCGCGCGCGCGCGCGCATCTTCAGAGCATGAC 1166
Db 969 CCAAGTGGCGGTGAAGATCGAGACAGTGAATTCAGAGACCATGCGCGCGCGCGCGCG 1028
Qy 1167 CAAGATCTGAGAGCGCTTCCGCGCGCGCGCAACCCGAGATCGATCTACAGAGCGCGCT 1226
Db 1029 CGAGAGAGCTGGCAGAGTACAGTCACTCTGTGTCGCCCGCGCTTTCGCGCGCGCT 1088
Qy 1227 GTACGTGGCAGCGACTTGAAGATCGCGAGCAACCGCGCGAGATGAGAGACTGGCGCA 1286
Db 1089 GGAAGGCAAGATCTGCAAGGCGCGGTGAGCGCGCGCGAGAGAGTGTCTTACTGGGCAAT 1148
Qy 1287 GCACTGCTGCGCGTGGGCTTACCAACCCCGAGCAAGAGCAACAGAGAGAGCGCGCTT 1346
Db 1149 CTGCGTGGCGCATGAGTGGCGCGCGCATGAAATATGCGCGCGCGCGCGCGCTGCGGA 1208
Qy 1347 CCGCGCATGAGCTGACCCCGAGCAAGTGAACCGTGAAGCGCGCGCATGAGCTGCGCGAGA 1406
Db 1209 CGCGAAGCAGACCGAATTCAGTCCGCGCGCGCGCATTCGATGATGATACCGA 1268
Qy 1407 GAGAGCTGAGACCGTGAACGACATCAAGAGCTGTGAGCAAGTGAACCTGAGCGCAAGCA 1466
Db 1269 GTGAGAGAGCGCGAGCGGACATCAAGCGCGAGCGAGGACCTCGAAGCTGAGCGCGCAC 1328

OY	1467	GATCATCCCGGCAATCAAGGTGGCGCACTGTGCAAGCTGTGTGGCGCGCGCAAGGCGCT	1526
Db	1329	CATGGCGCTGGGGCGGGCAAGCTCCGAAGTGGCTCCGGCACGCTGGCGCGCAGAAATTYA	1388
OY	1577	GACCGACATGTGTGCCCTCTGACCCGAGAGAGCGAGCTGAGGTGGCGCGGAACCGCGAGAT	1586
Db	1389	CGGCAACATCTGTCAACGAGCGCACCGCCACCGGTGACGAGCGCAACGTGAATCTATCGA	1448
OY	1587	CCTGCGCAGACCCGTGCAACGGCGTGTACTACAGACCCGACGAAG---GACTGTGATGGCCGA	1643
Db	1449	CCAGTCTCCGCAAGAGCGGGCGCTGTGTATCTCCGCGTCAACGACGCGCGCAACAGCTCACCGA	1508
OY	1644	GATCCAGAAGCAGAGGCCACAGACCAAGTGAACCTTACCAAGATTTAACGAGGCCCTTTCAGAA	1703
Db	1509	GATCTGTGAGCTGCGCGCGAGAGTTCATCCGTGTTCATCGGCGGTGCAAGTTCCACCCGA	1568
OY	1704	CCTGAGACCGGCAAGTAGTACGCC	1725
Db	1569	GTTCAAGTCCACCCCGTGGAC	1590

Search completed: December 31, 2005, 02:31:32
Job time : 8682.48 secs